

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: March 8, 2002, 09:22:17 ; Search time 10.85 Seconds
Maximum DB seq length: 0

Title: US-09-445-480a-2
Perfect score: 1287
Sequence: 1 MALSONQAKFSKGFVVMIW.....VDEGGVKVPPGSDVRFKRPTC 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 0
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_39;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1089	84.6	199	1 POST_ACTEQ	P81439 actinia equina
2	295.5	23.0	2766	1 THYG_BOVIN	P01267 bos taurus
3	295	22.9	2769	1 THYG_BOVIN	P06882 rattus norvegicus
4	294.5	22.9	2768	1 THYG_HUMAN	P01266 homo sapien
5	272	21.1	2768	1 THYG_HUMAN	Q11122 homo sapien
6	245.5	19.1	1375	1 NID2_HUMAN	P31226 rana catesbeiana
7	244	18.0	844	1 SAX_FRANCA	P08322 mus musculus
8	236	18.3	1403	1 NID2_MOUSE	P04441 mus musculus
9	156	12.1	279	1 HG2A_MOUSE	P10493 mus musculus
10	154	12.0	1245	1 NIDO_MOUSE	P10247 rattus norvegicus
11	153	11.9	280	1 HG2A_RAT	P14543 homo sapiens
12	153	11.9	1247	1 NIDO_HUMAN	P04233 homo sapiens
13	150	11.7	296	1 HG2A_HUMAN	Q05716 bos taurus
14	132.5	10.3	258	1 IBP4_BOVIN	P47879 mus musculus
15	129.5	10.1	254	1 IBP4_MOUSE	P21744 rattus norvegicus
16	128.5	10.0	254	1 IBP4_RAT	Q28939 ovis aries
17	124.5	9.7	237	1 IBP4_SHEEP	P22692 homo sapiens
18	122.5	9.5	258	1 IBP4_HUMAN	Q26985 sus scrofa
19	122.5	9.5	271	1 IBP5_PIG	P24593 homo sapiens
20	122.5	9.5	272	1 IBP5_HUMAN	P07079 mus musculus
21	120.5	9.4	271	1 IBP5_MOUSE	P24594 rattus norvegicus
22	118.5	9.2	271	1 IBP5_RAT	P24591 bos taurus
23	114	8.9	263	1 IBP1_BOVIN	P02469 mus musculus
24	114	8.9	1786	1 LMB1_MOUSE	P24592 homo sapiens
25	113.5	8.8	240	1 IBP6_HUMAN	P07942 homo sapiens
26	112	8.7	1786	1 LMB1_HUMAN	P16511 sus scrofa
27	110	8.5	266	1 IBP3_PIG	P20959 bos taurus
28	109.5	8.5	291	1 IBP1_HUMAN	P08833 homo sapiens
29	109.5	8.5	4544	1 LRPL_HUMAN	Q07754 homo sapiens
30	109.5	8.5	1 IBP3_MOUSE	P47879 mus musculus	
31	109	8.4	292	1 IBP3_RAT	P15473 rattus norvegicus
32	108.5	8.4	1	1 IBP6_RAT	P35572 rattus norvegicus

ALIGNMENTS					
RESULT	1	BOST_ACTEQ	STANDARD;	PRT;	199 AA.
ID	POST_ACTEQ				
AC	P81439;				
DT	15-DEC-1998 (Rel. 37, created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	EQUISPATIN.				
OS	Actinia equina (European sea anemone).				
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; zoantharia; Actiniaria;				
OC	Nyanthanthea; Actiniidae; Actinia.				
NCBI_TaxID	6106;				
RN					
RP					
RX					
RA					
RT					
RT					
RL					
RN					
RP					
RX					
RA					
RT					
RT					
J. Biol. Chem.	272:13899-13903(1997).				
[1]					
[2]					
SEQUENCE.					
REVISIONS.					
LENARCIĆ B., RITONJA A., STRUKELJ B., TURK B., TURK V.;					
J. Biol. Chem. 273:12682-12682(1998).					
-1 BIOL FUNCTION: POTENT INHIBITOR OF PAPAIN-LIKE CYSTEINE PROTEINASES					
AND HAS A PI OF 4.7.					
CC AND HAS A PI OF 4.7.					
CC -1 SIMILARITY: CONTAINS 3 THYROGLOBULIN TYPE-I DOMAINS.					
DR INTERPRO: IPR00716; THYROGLOBULIN_1.					
PFAM: PF00086; thyroglobulin_1; 3.					
SMART: SW0211; TY; 3.					
PROSITE: PS00484; THYROGLOBULIN_1; 2.					
KW THIOI PROTEASE INHIBITOR; REPEAT.					
FT DOMAIN 18 66 THYROGLOBULIN TYPE I 2.					
FT DOMAIN 86 134 THYROGLOBULIN TYPE I 2.					
FT DOMAIN 155 199 THYROGLOBULIN TYPE I 3.					
FT VARIANT 3 3 S -> T.					
FT VARIANT 42 42 C -> V.					
FT VARIANT 130 130 S -> T.					
FT VARIANT 137 137 P -> E.					
FT VARIANT 144 144 LQ -> IK.					
FT VARIANT 156 156 V -> L.					
FT VARIANT 184 184 G -> D.					
FT SEQUENCE 199 AA; 21755 MW; 87E44735F8171BFC CRC64;					

Query Match 84.6%; Score 1089; DB 1; Length 199;
Best Local Similarity 95.5%; Pred. No. 2.7e-90;
Matches 190; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 33 SLTKCQQLQASANSGLIGTYVPOCKEFEFEKQCMGSTGYCWCVDGKELTGKIRS 92
1 SLSKCQQLQASANSGLIGTYVPOCKEFEFEKQCMGSTGYCWCVDGKELTGKIRS 60

QY 93 PGDSRKALTLQMMQAIIVNPWGCGPSKEDGSFDEVQCCASNGECYCVDKGKEL 152
61 PGDSRKALTLQMMQAIIVNPWGCGPSKEDGSFDEVQCCASNGECYCVDKGKEL 120

FT	CONFLICT	1206	1206	S -> R (IN REF. 4)
SEQUENCE	2769 AA;	303218 MW;	1C7F227E9101DE2A	CRC64;
Qy	52 YVPPCCKERFEERKQCMGSGTYCWCYDGEIGKIRGSPPCSRRAKALIQMOMAI	111		
Db	48 YVPPCAKERFEEERKQCMGSGTYCWCYDGEIGKIRGSPPCSRRAKALIQMOMAI	102		
Qy	112 I-----WVPGWGMGPSSKADESFDDEVOCASNGECYCVDKGKELEGTRQQRPT-CER	165		
Db	103 ILLSYIYNTSTATSYLPQCDSDGDSYSPVQCDLRRQCVWCYDGEIGMVEYGRQQRPARPCR	162		
Qy	166 HLSCE-EARRKAHSNSRREVFMVPEGLDDSYNPKQC-----	202		
Db	163 ---SCEIRNRRLHGVDRSP---POCSPDAFRPVOCKLVNTTDMMIFLVHSYSRFPD	216		
Qy	203 -----WPS-TGICWCYDEGG 216			
Db	217 AFVTFSSFRSRFPEVSGCYCADSQ 242			
RESULT	4			
THY-G-RAT				
ID	-THYG-RAT			
AC	P06812; Q9JM94; Q9JKY6;		PRT;	2768 AA.
DT	01-JUN-1988 (Rel. 06, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	THYROGLOBULIN PRECURSOR.			
TG				
OS	Rattus norvegicus (Rat).			
OC	Bukayota; Metazoa; Chordata; Craniata; Vertebrate; Euteostomi; Rattus.			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBL-TAXID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WISTAR IMAMICHI;			
RA	Hishimura A.; Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=FISCHER 344;			
RA	Ding M., Jung C.-C., Cheng J.-M., Miyamoto T., Furudate S.I., Aguil T.;			
RT	"A missense mutation in the thyroglobulin gene causes hypothyroidism and dwarfism not associated with goiter in the WTC-rdw rat."			
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-213 FROM N.A.			
RC	STRAIN=FISCHER; TISSUE=Thymocytes;			
RX	MEDLINE=90220542; PubMed=2225666;			
RX	Graves P.N., Davies T.F.;			
RT	"A second thyroglobulin messenger RNA species (rgg-2) in rat thymocytes."			
RL	Mol. Endocrinol. 4:155-161(1990).			
RN	[4]			
RP	SEQUENCE OF 1-59 FROM N.A.			
RX	MEDLINE=86094383; PubMed=3555768;			
RA	Musti A.M., Avvedimento V.E., Polistina C., Ursini V.M., Obici S., Nitto L., Cocozza S., di Lauro R., Moscatelli C., Avvedimento V.E.;			
RT	"The complete structure of the rat thyroglobulin gene.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:323-327(1986).			
RN	[5]			
RP	SEQUENCE OF 1802-2768 FROM N.A.			
RX	MEDLINE=85154044; PubMed=33038512;			
RA	di Lauro R., Obici S., Condoluci D., Ursini V.M., Musti A.M.,			
RT	"The sequence of 967 amino acids at the carboxyl-end of rat thyroglobulin. Location and surroundings of two thyroxine-forming			
RT	sites";			
RL	Eur. J. Biochem. 148:7-11(1985);			
CC	(T4) AND TRITIOTODIYRONINE (T3).			
CC	- SUBUNIT: HOMODIMER.			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.			
CC	- PTM: SULFURATED (BY SIMILARITY).			
CC	- MISCELLANEOUS: IT IS NOT CERTAIN WHETHER THIS THYROGLOBULIN PLAYS ANY ROLE IN THE FUNCTION OF TRITIOTODIYRONINE.			
CC	- SIMILARITY: BELONGS TO THE TYPE B CARBOXYLESTERASE/LIPASE FAMILY.			
CC	- SIMILARITY: CONTAINS 11 THYROGLOBULIN TYPE-I DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on use by non profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements.html or send an email to license@isb-sib.ch).			
DR	EMBL: AB035201; BAA96132.1; -.			
DR	EMBL: AF21622; AAH34909.1; -.			
DR	EMBL: M3995; AAC2086.1; ALT TERM.			
DR	EMBL: M22558; AAH50379.1; JOINED.			
DR	EMBL: X0318; CA26183.1; -.			
DR	PIR: A22016; URT.			
DR	HSSP: P21836; 1MAH.			
DR	InterPro: IPR00776; Thyroglobulin_1.			
DR	pfam: PF00135; Coesterase_2.			
DR	SMART: SM00211; TY_10.			
DR	PROSITE: PS00484; THYROGLOBULIN_1; 9.			
DR	PROSITE: PS0041; CARBOXYLESTERASE_B_2; 1.			
FT	KW Glycoprotein; Repeat; Thyroid hormone; Iodination; Sulfation; Signal.			
FT	SIGNAL 1 REPEAT 20 POTENTIAL.			
FT	CHAIN 21 2768 SIGNAL 1			
FT	DOMAIN 32 93 THYROGLOBULIN.			
FT	DOMAIN 34 161 THYROGLOBULIN TYPE IA 1.			
FT	DOMAIN 162 298 THYROGLOBULIN TYPE IA 2.			
FT	DOMAIN 299 359 THYROGLOBULIN TYPE IA 3.			
FT	DOMAIN 605 658 THYROGLOBULIN TYPE IA 4.			
FT	DOMAIN 659 726 THYROGLOBULIN TYPE IA 5.			
FT	DOMAIN 727 922 THYROGLOBULIN TYPE IA 6.			
FT	DOMAIN 923 1074 THYROGLOBULIN TYPE IA 7.			
FT	DOMAIN 1075 1145 THYROGLOBULIN TYPE IA 8.			
FT	DOMAIN 1147 1211 THYROGLOBULIN TYPE IA 9.			
FT	REPEAT 1455 1468 TYPE II.			
FT	REPEAT 1469 1485 TYPE II.			
FT	REPEAT 1486 1502 TYPE II.			
FT	REPEAT 1501 1564 THYROGLOBULIN TYPE IB 2.			
FT	REPEAT 1602 1722 THYROGLOBULIN TYPE IB 1.			
FT	REPEAT 1723 1891 THYROGLOBULIN TYPE II A.			
FT	REPEAT 1892 1994 THYROGLOBULIN TYPE II A.			
FT	REPEAT 1995 2127 THYROGLOBULIN TYPE II B.			
FT	REPEAT 2128 2185 THYROGLOBULIN TYPE II A.			
FT	DISTRIFID 2265 2282 POTENTIAL.			
FT	MOD_RES 225 25 IODINATION (IN T4) (BY SIMILARITY).			
FT	MOD_RES 225 25 SULFATION (POTENTIAL).			
FT	MOD_RES 2574 2574 IODINATION (IN T4) (BY SIMILARITY).			
FT	MOD_RES 2588 2588 IODINATION (IN T4) (BY SIMILARITY).			
FT	MOD_RES 2766 2766 IODINATION (IN T3) (BY SIMILARITY).			
FT	CARBODYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 496 496 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 948 948 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBODYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).			

Db 1054 CHGKSDFCWCVDKDGRVQGT 1074
 RESULT 7
 SAX_RANCA STANDARD; PRT; 844 AA.
 ID SAX_RANCA
 AC P31226;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SAXIPHILIN PRECURSOR (SAX).
 OS Rana catesbeiana (Bull frog).
 OC Amphibia; Batrachia; Craniata; Vertebrata; Buteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
 OX NCBI_TaxID=8400;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94195765; PubMed=8146142;
 RA Morabito M.A.; Moczydłowski E.;
 RT "Molecular cloning of bullfrog saxiphilin: a unique relative of the
 transferrin family that binds saxitoxin";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:2478-2482(1994).
 RN [2]
 RP REVISIONS.
 RX MEDLINE=95327702; PubMed=7604048;
 RA Morabito M.A.; Moczydłowski E.;
 RL Proc. Natl. Acad. Sci. U.S.A. 92:6651-6651(1995).
 RN [3]
 RP SEQUENCE OF 20-44: 317-333; 360-371; 541-571; 606-624 AND 690-714.
 RC TISSUE=Plasma;
 RX MEDLINE=91332078; PubMed=1869567;
 RA Li Y.; Moczydłowski E.;
 RT "Purification and partial sequencing of saxiphilin, a
 RT saxitoxin-binding protein from the bullfrog, reveals homology to
 RT transferrin";
 RL J. Biol. Chem. 266:15481-15487(1991).
 CC -- FUNCTION: BINDS SPECIFICALLY TO THE NEUROTOXIN SAXITOXIN. ITS
 PHYSIOLOGICAL ROLE MAY BE TO TRANSPORT OR SQUEESE AN ENDOGENOUS
 ORGANIC MOLECULE OTHER THAN FE(3+). IT MAY PARTICIPATE IN A
 DETOXIFICATION MECHANISM FOR NEUTRALIZING A MICROBIAL TOXIN.
 CC -- SUBUNIT: MONOMER.
 CC -- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -- TISSUE SPECIFICITY: PLASMA, HIGHEST LEVELS OF TRANSCRIPTS FOUND IN
 THE LIVER, THE LUNG, THE PANCREAS AND THE BRAIN.
 CC -- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.
 CC -- SIMILARITY: RELATED TO TRANSFERRIN (BUT DOES NOT BIND FE(3+)).
 CC -- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC --
 DR EMBL; U05246; AAA75440.1; .
 DR PIR; A39426;
 DR HSSP; P56410; 1AOY
 DR InterPro; IPR00116; Thyroglobulin_1.
 DR InterPro; IPR00156; Transferrin.
 DR Pfam; PRO0086; thyroglobulin_1; 2.
 DR Pfam; PRO0045; transferrin; 4.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SM00034; TY_FER; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 KW Signal; Extracellular matrix; Repeat.
 FT SIGNAL 1
 FT CHAIN 20 84 SAXIPHILIN.
 FT DOMAIN 124 172 THYROGLOBULIN TYPE I 1.
 FT DOMAIN 196 244 THYROGLOBULIN TYPE I 2.

FT DOMAIN 109 251 ABSENT IN TRANSFERRINS.
 FT DISULFID 29 64 BY SIMILARITY.
 FT DISULFID 39 55 BY SIMILARITY.
 FT DISULFID 277 360 BY SIMILARITY.
 FT DISULFID 322 335 BY SIMILARITY.
 FT DISULFID 332 343 BY SIMILARITY.
 FT DISULFID 388 402 BY SIMILARITY.
 FT DISULFID 495 527 BY SIMILARITY.
 FT DISULFID 505 518 BY SIMILARITY.
 FT DISULFID 552 839 BY SIMILARITY.
 FT DISULFID 570 799 BY SIMILARITY.
 FT DISULFID 607 685 BY SIMILARITY.
 FT DISULFID 641 655 BY SIMILARITY.
 FT DISULFID 652 668 BY SIMILARITY.
 FT DISULFID 725 739 BY SIMILARITY.
 SQ SEQUENCE 844 AA; 93089 MW; 162A765AD02C3C5E CRC64;

Query Match 19.0%; Score 244; DB 1; Length 844;
 Best Local Similarity 28.9%; Pred. No. 2.4e-14;
 Matches 57; Conservative 25; Mismatches 47; Indels 68; Gaps 5;

Qy 29 STEASLTKC-QQOQASANSGLITGVYWPQCKTGEFEKEEKQCGSGTYCYCVCYDGDKEKGIT 87
 Db 102 SSNRDQCLKLERQOQALAKKMIGHYIPOQDGKQHGNQPPQOCHGSGTGHCVNAMGEKISGT 161
 Qy 88 KIRGSFDCSRRAKALTLCOMMOAITIVNPGRGPGPSCKADGSFDEVOCCASNGECYCVD 147
 Db 162 NT-----PPG----- 166
 Qy 148 KCKELEGTRQGRGTCERH-LSEC-BEARIAHNSLRLRMEVPECELEDGSYNNPVQWPS 205
 Db 167 -----QTRMCEHPLKLKRQVALGGDEKVLGFVPOCDERKGNYEPQOOGHS 216
 Qy 206 TGYCWVDEGGVKVPGS 222
 Db 217 TGYSWCVNAEGEEATG 233

RESULT 8
 NID2_MOUSE STANDARD; PRT; 1403 AA.
 ID NID2_MOUSE
 AC O98322;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE NIDGEN-2 PRECURSOR (NID-2) (ENACTIN-2).
 GN NID2.
 OS Mus musculus (Mouse).
 OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95297167; PubMed=9533511;
 RA Kimura N.; Toyoshima T.; Kojima T.; Shimane M.;
 RT Enactin-2: a new member of basement membrane protein with high
 RT homology to enactin/nidogen.;
 RL Exp. Cell Res. 241:36-45 (1998);
 CC --
 CC -- FUNCTION: CELL ADHESION GLICOPROTEIN. MIGHT BE INVOLVED IN CELL-
 CC -- OSTEOBLAST DIFFERENTIATION. IT PROBABLY HAS A ROLE IN CELL-
 CC -- EXTRACELLULAR MATRIX INTERACTIONS.
 CC -- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
 CC -- SIMILARITY: HIGHLY N- AND O-GLUCOSYLATED (BY SIMILARITY).
 CC -- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.
 CC -- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
 CC --
 CC -- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC

DR EMBL X05128; CAA29010_1; :-;
 DR EMBL X05129; CAB29012_1; :-;
 DR EMBL X05420; CAB31297_1; :-;
 DR EMBL X13414; :- NOT_ANNOTATED_CDS.
 DR PIR; A02244; H-HMSG.
 DR PIR; A27866; A27866.
 DR PIR; B27866; B27866.
 DR PIR; S03099; S03099.
 DR PIR; S0284; S0284.
 DR HSSP; P04233; IAGA.
 DR MGD; MGI-96534; II.
 DR InterPro; IPRO00716; Thyroglobulin_1.
 DR Pfam; PF00086; thyroglobulin_1.
 DR SMART; SM00211; TY; 1.
 DR PROSITE; PS00444; THYROGLOBULIN_1; 1.
 KW Glycoprotein; MHC; Proteoglycan; Transmembrane; Alternative splicing;
 KW Signal-anchor.
 FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL) MEMBRANE PROTEIN
 FT TRANSMEM 30 55 SIGNAL-ANCHOR (TYPE-II) MEMBRANE PROTEIN
 FT DOMAIN 56 279 (POTENTIAL).
 FT CARBOIID 113 113 EXTRACELLULAR (POTENTIAL).
 FT CARBOIID 119 119 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOIID 265 265 O-LINKED (GLUCOSAMINOGLYCAN).
 FT DOMAIN 211 254 THYROGLOBULIN TYPE I.
 FT VARSPLIC 192 255 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 10 10 N -> M (IN REF.).
 FT CONFLICT 113 113 N -> Q (IN REF.).
 FT CONFLICT 229 231 STG -> RHC (IN REF.).
 SQ 279 AA; 31557 MW; 30085014 FI70752 CRG64;

Query Match 12.1%; Score 156; DB 1; Length 279;
 Best Local Similarity 40.7%; Pred. No. 5.6e-0%;
 Matches 35; Conservative 11; Mismatches 32; Indels 8; Gaps 2;

Qy 19 WVLFACAIT----STEAS--LTTCQQLQASANSGLIGTYVPOCKEIGEEFEERQCWGS 70
 Db 170 WILFEMSKNSLLEKKPTEAFFPKVLCQEVSHIPAVVYGAFLRKCDENGNYLPLQCHGS 229

Qy 71 TGYCMWDEGGKELGTKRRSPDCS 96
 Db 230 TGYCMWCEVPGTETEVPHTKSRGRHNCs 255

RESULT 10

ID NIDO_MOUSE STANDARD; PRT; 1245 AA.

AC P10493; 1-149 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE NIDOGEN PRECURSOR (ENTACTIN).

GN NID OR NDI OR ENT.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.

RY MEDLINE=830791005; NCBI_TaxId=10090; RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 29-40.

RA Durkin M.E.; Chakravarti S.; Bartos B.B.; Liu S.H.; Friedman R.L.; Chung A.E.; Mann K.; Deutermann R.; Aumailley M.; Timpl R.; Raimondi L.; Yamada Y.; Pan T.-C.; Conway D.; Chu M.-L.; "Amino acid sequence and domain structure of entactin. Homology with epidermal growth factor precursor and low density lipoprotein receptor.", J. Cell Biol. 107:2749-2756(1988). [2]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=82231638; PubMed=2456973;

RA Mann K.; Deutermann R.; Aumailley M.; Timpl R.; Raimondi L.; Chung A.E.; "Amino acid sequence and domain structure of entactin. Homology with epidermal growth factor precursor and low density lipoprotein receptor.", J. Cell Biol. 107:2749-2756(1988).

RN [2]

RP CarBBank; CCSB; 37018; :-.

DR CarBBank; CCSB; 37019; :-.

DR MGD; MGI-97342; Nid1.

DR InterPro; IPRO00152; Ask_hydroxyl.

DR InterPro; IPRO00561; EGF-like.

DR InterPro; IPRO01881; EGFR_Ca.

DR InterPro; IPRO00033; Ldl_receptor_rep.

DR InterPro; IPRO03886; Nidogen_ext.

DR InterPro; IPRO00716; Thyroglobulin_1.

DR Pfam; PF00008; EGF; 6.

DR Pfam; PF00058; ldl_recept_b; 3.

DR Pfam; PF00086; thyroglobulin_1.

RT membrane protein with binding activity for laminin, collagen IV and cells.;

RT EMBL X05129; CAB29012_1; :-;

RT EMBL X05420; CAB31297_1; :-;

RT EMBL X13414; :- NOT_ANNOTATED_CDS.

RN [3]

RN SEQUENCE OF 1-251 FROM N.A.

RC STRAIN=BALB/C; TISSUE=Liver;

RX MEDLINE=94040711; PubMed=8224873;

RA Durkin M.E.; Liou S.H.; Reing J.; Chung A.E.; "Characterization of the 5' end of the mouse Ent gene encoding the basement membrane protein, entactin.", Gene 132:261-266(1993).

RN [4]

RN SEQUENCE OF 1207-1245 FROM N.A.

RC STRAIN=C57BL/6J x CBA/J;

RX MEDLINE=93324912; PubMed=7601446;

RA Durkin M.E.; Newer U.M.; Chung A.E.; "Exon organization of the mouse entactin gene corresponds to the structural domains of the polypeptide and has regional homology to the low-density lipoprotein receptor gene.", Genomics 26:219-228(1995).

RN [5]

RA Paulsson M.; Deutzmann R.; Dziadek M.; Nowack H.; Timpl R.; Weber S.; Engel J.; "Purification and structural characterization of intact and fragmented nidogen obtained from a tumor basement membrane.", Eur. J. Biochem. 156:467-478(1986).

RN [6]

RA Carbhydrate-linkage sites, and partial sequence.

RX MEDLINE=93316903; Pubmed=8346911;

RA Fujikawa S.; Shinkai H.; Mann K.; Timpl R.; "Structure and localization of O- and N-linked oligosaccharide chains on basement membrane protein nidogen.", Matrix 13:215-222(1993).

CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ; ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.

CC -!- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.

CC -!- PTM: N- AND O-GLYCOSEYLATION.

CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstations on the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annou> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X14194; CAA2408_1; :-.

DR EMBL; X14364_1; :-.

DR EMBL; L11324; AAA77652_1; :-.

DR EMBL; L17322; AAA77652_1; JOINED.

DR EMBL; L17323; AAA77655_1; JOINED.

DR EMBL; X83093; CAAS0148_1; :-.

DR PIR; S02230; MMSND.

DR HSSP; P07204; 2ADX.

DR CarrBBank; CCSB; 37018; :-.

DR CarrBBank; CCSB; 37019; :-.

DR MGD; MGI-97342; Nid1.

DR InterPro; IPRO00152; Ask_hydroxyl.

DR InterPro; IPRO00561; EGF-like.

DR InterPro; IPRO01881; EGFR_Ca.

DR InterPro; IPRO00033; Ldl_receptor_rep.

DR InterPro; IPRO03886; Nidogen_ext.

DR InterPro; IPRO00716; Thyroglobulin_1.

DR Pfam; PF00008; EGF; 6.

DR Pfam; PF00058; ldl_recept_b; 3.

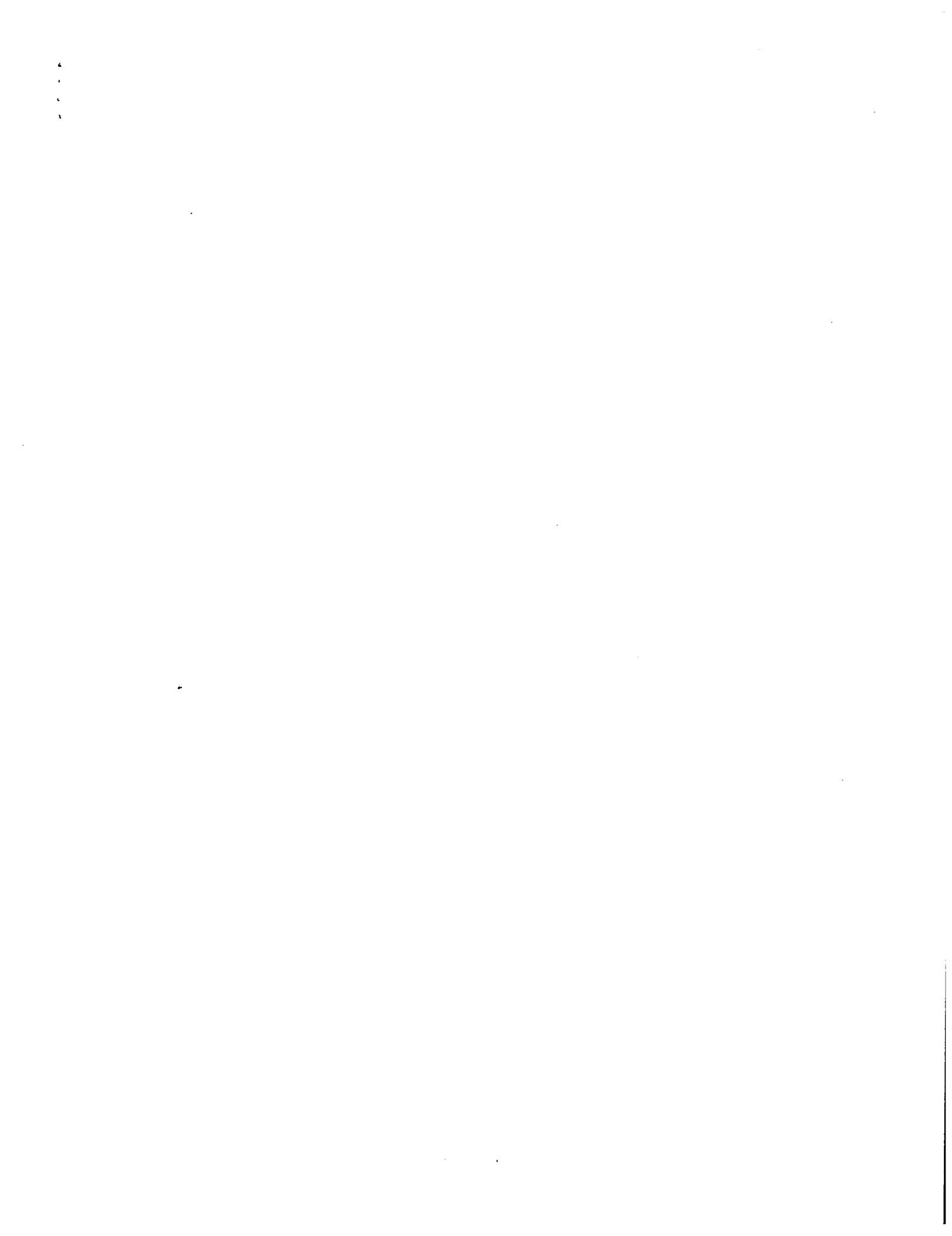
DR Pfam; PF00086; thyroglobulin_1.

FT DOMAIN	57	280	EXTRACELLULAR (POTENTIAL).	CC
FT CARBOHYD	114	114	N-LINKED (GlcNAc, . .) (POTENTIAL).	CC
FT CARBOHYD	120	120	O-LINKED (GlcNAc, . .) (POTENTIAL).	CC
FT CARBOHYD	266	266	SIMILARITY; MISSING (IN SHORT ISOFORM).	CC
FT DOMAIN	212	255	THYROGLOBULIN TYPE I.	DR
FT VARSPIC	193	256	MISSING (IN SHORT ISOFORM).	DR
SQ SEQUENCE	280 AA;	31642 MW;	D935D169A9B85732 CRC64;	DR
Query Match		11.9%	Score 153; DB 1; Length 280;	
Best Local Similarity	38.4%	Pred.	No. 1e-06;	
Matches	33;	Conservative	10;	Mismatches
OY	19	WVLFATCATISTEAS-.-.-.-.LTQKQQLQASANGLIGTYVPOCKEGEREKOWGS	70	8; Gaps 1;
Db	171	WLFEMNSKNSLLEKQPTQTPPKVLUKQEEWSHIPDWHGAFRPRCDENGYNPLQCGS	230	
OY	71	TGTCNCVQDVDEGGKEILGTKIRGSPDCS	96	
Db	231	TGTCNCVQFPNGTEVPHIKSRRGRHNCs	256	
RESULT 12				
NID0_HUMAN				
ID NID0_HUMAN STANDARD;			PRM; 1247 AA.	
AC P14543; O14942;				
DT 01-JAN-1990 (Rel. 13, created)				
DT 01-JAN-1990 (Rel. 13, last sequence update)				
DT 20-AUG-2001 (Rel. 40, last annotation update)				
DE NITROGEN PRECURSOR (ENTACTIN).				
GN NID.				
OS Homo sapiens (Human).				
OC Hominidae; Euteleostomi;				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; HOMO.				
OX NCBI_TAXID=9606;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=90091745; PubMed=2574658;				
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,				
RA Chu M.-L., Knowton R., Deutzmann R., Timpl R., Dittlo J.;				
RT "Human nidogen: complete amino acid sequence and structural domains deduced from cDNAs, and evidence for polymorphism of the gene";				
RT DNA 8:581-594 (1989).				
RN [2]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=9604428; PubMed=7557988;				
RA Zimmermann K., Houschen S., Hafer M., Nischk R.;				
RT "Genomic sequences and structural organization of the human nidogen gene (NID)."; Genomics 27:245-250(1995).				
RN [3]				
RP SEQUENCE OF 667-1247 FROM N.A.				
RC TISSUE=Placenta;				
RX MEDLINE=832270475; PubMed=2471403;				
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,				
RA Timpl R., Chu M.-L., Dittlo J.;				
RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1q43.>";				
RL Am. J. Hum. Genet. 44:876-885 (1989).				
CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.				
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.				
CC -!- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.				
CC -!- PIM: N- AND O-GLYCOSYLATED.				
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.				
CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.				
CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).				
CC DR EMBL; M30269; CA059932.1; -.				
CC DR EMBL; X82245; CA057709.1; -.				
CC DR EMBL; X84819; CA057709.1; JOINED.				
CC DR EMBL; X84820; CA057709.1; JOINED.				
CC DR EMBL; X84821; CA057709.1; JOINED.				
CC DR EMBL; X84822; CA057709.1; JOINED.				
CC DR EMBL; X84823; CA057709.1; JOINED.				
CC DR EMBL; X84824; CA057709.1; JOINED.				
CC DR EMBL; X84825; CA057709.1; JOINED.				
CC DR EMBL; X84826; CA057709.1; JOINED.				
CC DR EMBL; X84827; CA057709.1; JOINED.				
CC DR EMBL; X84828; CA057709.1; JOINED.				
CC DR EMBL; X84829; CA057709.1; JOINED.				
CC DR EMBL; X84830; CA057709.1; JOINED.				
CC DR EMBL; X84831; CA057709.1; JOINED.				
CC DR EMBL; X84832; CA057709.1; JOINED.				
CC DR EMBL; X84833; CA057709.1; JOINED.				
CC DR EMBL; X84834; CA057709.1; JOINED.				
CC DR EMBL; X84835; CA057709.1; JOINED.				
CC DR EMBL; X84836; CA057709.1; JOINED.				
CC DR EMBL; X84837; CA057709.1; JOINED.				
CC DR EMBL; M27445; AAA57761.1; -.				
CC DR IR; A3332; MMUND.				
CC DR HSSP; P07204; 2ADX.				
CC DR MIM; 131390; -.				
CC DR InterPro; IPR000152; ASX hydroxyl.				
CC DR InterPro; IPR00561; EGF like.				
CC DR InterPro; IPR00770; EGF_Ca.				
CC DR InterPro; IPR00033; Ldl_receptor_rep.				
CC DR SMART; M00386; Nidogen_ext.				
CC DR SMART; M00386; Thyroglobulin_1.				
CC DR SMART; SMO0135; LY; 5.				
CC DR SMART; SMO0339; NID0; 1.				
CC DR SMART; SMO0211; TY; 1.				
CC DR PROSITE; PS00010; ASX_HYDROXYL; 3.				
CC DR PROSITE; PS00001; EGF_LIKE; 2.				
CC DR PROSITE; PS00022; EGFI; 1.				
CC DR PROSITE; PS01186; EGFI_2; 5.				
CC DR PROSITE; PS0187; EGF_CA; 2.				
CC KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation; Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.				
CC FT SIGNAL 1 28				
CC FT CHAIN 29 1247 NID0.				
CC FT DOMAIN 29 659 I (LARGER GLOBULAR DOMAIN).				
CC FT DOMAIN 670 917 I (CYSTEINE-RICH).				
CC FT DOMAIN 918 1247 I (SMALLER GLOBULAR DOMAIN).				
CC RA Timpl R., Chu M.-L., Dittlo J.;				
CC RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1q43.";				
CC RL Am. J. Hum. Genet. 44:876-885 (1989).				
CC CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.				
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.				
CC CC -!- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.				
CC CC -!- PIM: N- AND O-GLYCOSYLATED.				
CC CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.				
CC CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.				
CC CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.				
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).				
CC DR EMBL; M30269; AA059932.1; -.				
CC DR EMBL; X82245; CA057709.1; -.				
CC DR EMBL; X84819; CA057709.1; JOINED.				
CC DR EMBL; X84820; CA057709.1; JOINED.				
CC DR EMBL; X84821; CA057709.1; JOINED.				
CC DR EMBL; X84822; CA057709.1; JOINED.				
CC DR EMBL; X84823; CA057709.1; JOINED.				
CC DR EMBL; X84824; CA057709.1; JOINED.				
CC DR EMBL; X84825; CA057709.1; JOINED.				
CC DR EMBL; X84826; CA057709.1; JOINED.				
CC DR EMBL; X84827; CA057709.1; JOINED.				
CC DR EMBL; X84828; CA057709.1; JOINED.				
CC DR EMBL; X84829; CA057709.1; JOINED.				
CC DR EMBL; X84830; CA057709.1; JOINED.				
CC DR EMBL; X84831; CA057709.1; JOINED.				
CC DR EMBL; X84832; CA057709.1; JOINED.				
CC DR EMBL; X84833; CA057709.1; JOINED.				
CC DR EMBL; X84834; CA057709.1; JOINED.				
CC DR EMBL; X84835; CA057709.1; JOINED.				
CC DR EMBL; X84836; CA057709.1; JOINED.				
CC DR EMBL; X84837; CA057709.1; JOINED.				
CC DR EMBL; M27445; AAA57761.1; -.				
CC DR IR; A3332; MMUND.				
CC DR HSSP; P07204; 2ADX.				
CC DR MIM; 131390; -.				
CC DR InterPro; IPR000152; ASX hydroxyl.				
CC DR InterPro; IPR00561; EGF like.				
CC DR InterPro; IPR00770; EGF_Ca.				
CC DR SMART; SMO0179; EGF_CA; 2.				
CC DR SMART; SMO0001; EGF_like; 4.				
CC DR SMART; SMO0135; LY; 5.				
CC DR SMART; SMO0339; NID0; 1.				
CC DR SMART; SMO0211; TY; 1.				
CC DR PROSITE; PS00010; ASX_HYDROXYL; 3.				
CC DR PROSITE; PS00001; EGF_LIKE; 2.				
CC DR PROSITE; PS00022; EGFI; 1.				
CC DR PROSITE; PS01186; EGFI_2; 5.				
CC DR PROSITE; PS0187; EGF_CA; 2.				
CC KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation; Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.				
CC FT SIGNAL 1 28				
CC FT CHAIN 29 1247 NID0.				
CC FT DOMAIN 29 659 I (LARGER GLOBULAR DOMAIN).				
CC FT DOMAIN 670 917 I (CYSTEINE-RICH).				
CC FT DOMAIN 918 1247 I (SMALLER GLOBULAR DOMAIN).				
CC RA Timpl R., Chu M.-L., Dittlo J.;				
CC RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1q43.";				
CC RL Am. J. Hum. Genet. 44:876-885 (1989).				
CC CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.				
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.				
CC CC -!- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.				
CC CC -!- PIM: N- AND O-GLYCOSYLATED.				
CC CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.				
CC CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.				
CC CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.				
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).				
CC DR EMBL; M30269; AA059932.1; -.				
CC DR EMBL; X82245; CA057709.1; -.				
CC DR EMBL; X84819; CA057709.1; JOINED.				
CC DR EMBL; X84820; CA057709.1; JOINED.				
CC DR EMBL; X84821; CA057709.1; JOINED.				
CC DR EMBL; X84822; CA057709.1; JOINED.				
CC DR EMBL; X84823; CA057709.1; JOINED.				
CC DR EMBL; X84824; CA057709.1; JOINED.				
CC DR EMBL; X84825; CA057709.1; JOINED.				
CC DR EMBL; X84826; CA057709.1; JOINED.				
CC DR EMBL; X84827; CA057709.1; JOINED.				
CC DR EMBL; X84828; CA057709.1; JOINED.				
CC DR EMBL; X84829; CA057709.1; JOINED.				
CC DR EMBL; X84830; CA057709.1; JOINED.				
CC DR EMBL; X84831; CA057709.1; JOINED.				
CC DR EMBL; X84832; CA057709.1; JOINED.				
CC DR EMBL; X84833; CA057709.1; JOINED.				
CC DR EMBL; X84834; CA057709.1; JOINED.				
CC DR EMBL; X84835; CA057709.1; JOINED.				
CC DR EMBL; X84836; CA057709.1; JOINED.				
CC DR EMBL; X84837; CA057709.1; JOINED.				
CC DR EMBL; M27445; AAA57761.1; -.				
CC DR IR; A3332; MMUND.				
CC DR HSSP; P07204; 2ADX.				
CC DR MIM; 131390; -.				
CC DR InterPro; IPR000152; ASX hydroxyl.				
CC DR InterPro; IPR00561; EGF like.				
CC DR InterPro; IPR00770; EGF_Ca.				
CC DR SMART; SMO0179; EGF_CA; 2.				
CC DR SMART; SMO0001; EGF_like; 4.				
CC DR SMART; SMO0135; LY; 5.				
CC DR SMART; SMO0339; NID0; 1.				
CC DR SMART; SMO0211; TY; 1.				
CC DR PROSITE; PS00010; ASX_HYDROXYL; 3.				
CC DR PROSITE; PS00001; EGF_LIKE; 2.				
CC DR PROSITE; PS00022; EGFI; 1.				
CC DR PROSITE; PS01186; EGFI_2; 5.				
CC DR PROSITE; PS0187; EGF_CA; 2.				
CC KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation; Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.				
CC FT SIGNAL 1 28				
CC FT CHAIN 29 1247 NID0.				
CC FT DOMAIN 29 659 I (LARGER GLOBULAR DOMAIN).				
CC FT DOMAIN 668 709 I (CYSTEINE-RICH).				
CC RA Timpl R., Chu M.-L., Dittlo J.;				
CC RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1q43.";				
CC RL Am. J. Hum. Genet. 44:876-885 (1989).				
CC CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.				
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.				
CC CC -!- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.				
CC CC -!- PIM: N- AND O-GLYCOSYLATED.				
CC CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.				
CC CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.				
CC CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.				
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).				
CC DR EMBL; M30269; AA059932.1; -.				
CC DR EMBL; X82245; CA057709.1; -.				
CC DR EMBL; X84819; CA057709.1; JOINED.				
CC DR EMBL; X84820; CA057709.1; JOINED.				
CC DR EMBL; X84821; CA057709.1; JOINED.				
CC DR EMBL; X84822; CA057709.1; JOINED.				
CC DR EMBL; X84823; CA057				

FT	DISULFID	721	736	BY SIMILARITY.	RT	"cDNA clone for the human invariant gamma chain of class II histocompatibility antigens and its implications for the protein
FT	DISULFID	738	750	BY SIMILARITY.	RT	histocompatibility antigens and its implications for the protein
FT	DISULFID	752	777	BY SIMILARITY.	RT	structure.";
FT	DISULFID	769	787	BY SIMILARITY.	RT	Proc. Natl. Acad. Sci. U.S.A. 80:7395-7399(1983).
FT	DISULFID	789	800	BY SIMILARITY.	RN	[4]
FT	DISULFID	806	817	BY SIMILARITY.	RP	SEQUENCE FROM N.A.
FT	DISULFID	811	826	BY SIMILARITY.	RX	Medline-86233451; PubMed=3459184;
RA	Strubin M., Mach B., Long E.O.;	828	839	BY SIMILARITY.	RA	O'Sullivan D.M., Lachammar D., Wilson M.C., Peterson P.A.,
RT	The complete sequence of the mRNA for the HLA-DR-associated class II histocompatibility antigens.;"	1212	1223	BY SIMILARITY.	RA	Quaranta V.;
RT	Nucleic Acids Res. 13:8827-8841(1985).	1219	1232	BY SIMILARITY.	RT	"Structure of the human Ia-associated invariant (gamma)-chain gene: identification of 5' sequences shared with major histocompatibility complex class II genes.";
FT	DISULFID	1234	1243	BY SIMILARITY.	RT	complex class II genes.";
FT	SITE	1702	704	CELL ATTACHMENT SITE.	RL	Proc. Natl. Acad. Sci. U.S.A. 83:4484-4488(1986).
FT	CARBOHYD	1137	1137	N-LINKED (GLCNAC. . .) (POTENTIAL).	RN	[5]
FT	CONFLICT	33	34	EL-> SS (IN REF. 2).	RP	X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS) OF 103-117.
FT	CONFLICT	37	42	FGP3G -> SAADR (IN REF. 2).	RX	Medline-9605023; PubMed=747400;
FT	SEQUENCE	1115	1115	H -> T (IN REF. 3).	RA	Ghosh P., Amaya M., Mellins E., Wiley D.C.;
SO		1247	AA:	136488 MW; 4681B3C1575B CRC64;	RA	"The structure of an intermediate in class II MHC maturation: CLIP bound to H-DR3.";
				RL Nature 378:457-462(1995).	RL	bound to H-DR3.";
				[6]	RN	
Query	Match		11.9%	Score 153; DB 1; Length 1247;	RX	STRUCTURE BY NMR OF 134-208.
Best	Local Similarity		25.5%	Pred. NO. 4.5e-06;	RX	Medline-99059718; PubMed=9843686;
Matches	56;	Conservative	18;	Mismatches 70; Indels 76; Gaps 11;	RT	"Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein 11.";
Db	750	CAITSEASLTKCQOLQASANSLGLTGVPP--OCKEGEPEEKQKGSTEGCWV--D	78	750 CAVAVDQDPINCE-----TGHLNDCHFDIPORAQCIVY-----GSSTYCSCLPGFS	794	EMBO J. 17:5812-6818(1998).
Qy	79	ERKEKEIGLKIRGSFDCSRRAKALTICQMQAMQATIVNGWCGPPSKADGSEDEVOCAS	138	795 GDQ-----ACQDVDE-----COPSRCHPDAF----CYNT	820	CC -!- FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE PEPTIDE LOADING OF CLASS II TAKES PLACE.
Qy	139	NEGCYCVDKKKELEGRQGRPTC--EHLLSCEERIKAH-----SNSLRVE	185	Db 821 PGSFPTCQCKPCYOGGFR-----CVPGYEVKTRCQHR--EHILGAAGATDPQRPIP	872	CC -!- SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETERO TRIMERS.
Db	873	LFPVPECDAHGHYAPQOCHGSTGTYCWCVRDGREGREVEGTRTR	912	Qy 186 MIVPECEDEDGQNPWGPSTGVCWVDRGGVKPGSIVR	225	CC -!- SIMILARITY: CONTAINS 1 THYROID GLOBULIN TYPE-I DOMAIN.
Db				CC WWW="http://www.ncbi.nlm.nih.gov/prowt/cd/cd4.htm".	CC	-!- DATABASE: NAME=PROW; NOTE=CD entry CD74 entry;
RESULT	13			CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	CC	
ID	HGCA_HUMAN	STANDARD;	PRT;	CC	CC	
AC	PP4233; Q28332;		296 AA.	CC	CC	
DT	20-MAR-1987 (Rel. 04, Created)			CC	CC	
DT	01-JUL-1993 (Rel. 26, Last sequence update)			CC	CC	
DT	20-AUG-2001 (Rel. 40, Last annotation update)			CC	CC	
DE	HLA CLASS II HISTOCOMPATIBILITY ASSOCIATED INVARIANT CHAIN (P33) (CD74 ANTIGEN).			CC	CC	
DE	ASSOCIATED INVARIANT CHAIN (P33) (CD74 ANTIGEN).			CC	CC	
GN	C074 OR DHM4.			CC	CC	
OS	Homo sapiens (Human).			CC	CC	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			CC	CC	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			CC	CC	
OX	NCBI_TaxID=9606;			CC	CC	
RN	[1]			CC	CC	
RP	SEQUENCE FROM N.A.			CC	CC	
RX	MEDLINE-86033681; PubMed=3001652;			CC	CC	
RA	Kudo J., Chao L.-Y., Narni F., Saunders G.F.;			CC	CC	
RT	"Structure of the human gene encoding the invariant gamma-chain of class II histocompatibility antigens."			CC	CC	
RL	Invariant chain reveals a polypeptide with an unusual transmembrane polarity.";			CC	CC	
RT	RT invariant chain reveals a polypeptide with an unusual transmembrane polarity.";			CC	CC	
RL	Embo J. 3:859-872(1984).			CC	CC	
RN	[3]			CC	CC	
RP	SEQUENCE OF 27-96 FROM N.A.			CC	CC	
RX	MEDLINE-84170234; PubMed=6324166;			CC	CC	
RA	Claesson L., Larhammar D., Rask L., Peterson P.A.;			CC	CC	

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X81882; CAA57272; 1; -.
 DR EMBL; X76066; CAA53667; 1; -.
 DR EMBL; Z95492; CAB08959; 1; -.
 DR HSSP; P01180; INPO.
 DR MGI; MGI:96439; Igfp4.
 DR InterPro; IPR000867; IGF_BINDING.
 DR InterPro; IPR000716; Thyroglobulin_1.
 DR Pfam; PF00219; Igfbp; 1.
 DR Pfam; PF0086; thyroglobulin_1; 1.
 DR SMART; SM00121; IB; 1.
 DR SMART; SM00211; TY; 1.
 DR PROSITE; PS00222; IGF_BINDING; 1.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
 KW Growth factor binding; signal; Glycoprotein.
 FT SIGNAL 1 21 BY SIMILARITY.
 FT CHAIN 22 234 INSULIN-LIKE GROWTH FACTOR BINDING
 FT PROTEIN 4.
 DOMAIN 196 245 THYROGLOBULIN TYPE I.
 FT CARBOHYD 125 125 N LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 4 5 FG > CS (IN REF. 1).
 FT CONFLICT 13 13 A > T (IN REF. 1).
 FT CONFLICT 56 56 C > S (IN REF. 1).
 FT CONFLICT 67 67 C > G (IN REF. 1).
 FT CONFLICT 254 AA; 27807 MW; 58EF89CB514AEE17 CRC64;
 SQ -----
 Query Match 10.1%; Score 129.5; DB: 1; Length 254;
 Best Local Similarity 24.2%; Pred. No. 0.0012; Indels 53; Gaps 12;
 Matches 52; Conservative 34; Mismatches 76; Gaps 12;
 OY 37 COOLQASANGL-----IGIVPOCKETGEFEKOCWGSGTGWCWCVDEDGKEIL 85
 Db 44 CEELVRPGGCCATCALGLGMPCGVTPRG-----SGMCYPPIKG-----EKPL 90
 OY 86 GTKTRGSPDCSRRAALTLCOMMQLITVNPWGCGPPCKADGSFDEVQCCASNGECY-- 143
 Db 91 RTLMHGQCVC-----TELSEIAQESLOQTSDKDESEHPRNSFN--PCSAHDHRCLQK 141
 OY 144 ---CVDIKGKELEG-FRQGRP---TCERILSECERAKTAHSNLSLRVEMF-- VPEC 191
 Db 142 HMAKIRDNSKMKIVGTPTPREEPVPVPGSCOSEHLRALE-RUASQSRTHEDOLFIPIN 200
 OY 192 LEDGSYNNVQCWPS---TGYCMVCD-BGGVKVPG 221
 Db 201 DRNGNFHKOCHFALDGDGRGKWCVDRTGKVLPK 235

Search completed: March 8, 2002, 09:25:29
 Job time: 192 sec



GeCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:24:02 ; Search time 23.53 Seconds
(without alignments) Million cell updates/sec
(without alignments) Million cell updates/sec
1435.993

Title: US-09-445-480a-2
Perfect score: 1287
Sequence: 1 MALSQNQAKFSKGKVMMIWV... ... VDEGGVKVPGSDVFRKRPC 231

Scoring table: BLOSUM62
Gapext 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRIMBL17:
 1: sp_archaea:
 2: sp_bacteria:
 3: sp_fungi:
 4: sp_human:
 5: sp_invertebrate:
 6: sp_mammal:
 7: sp_mhc:
 8: sp_organelle:
 9: sp_phage:
 10: sp_plant:
 11: sp_rabbit:
 12: sp_virus:
 13: sp_vertebrate:
 14: sp_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1287	100.0	231	5 Q9u6KB
2	1278	99.3	231	5 Q9u4RB
3	1279	21.7	212	6 Q9e3SO
4	259.5	20.2	446	11 Q9ER95
5	259.5	20.1	457	4 Q91300
6	255.5	19.9	452	11 Q9WVN9
7	252.5	19.6	434	4 Q9H4FB
8	252.5	19.6	445	9 Q913U7
9	238.5	18.5	116	5 Q94901
10	177.5	13.8	188	13 Q98UBD
11	173	613	5 Q9Y5F6	Q9y5f6 drosophila
12	171.5	13.3	1511	5 Q9VB21
13	167.5	13.0	126	4 Q9ICV2
14	167	13.0	234	7 Q9RE4
15	157	13.0	237	13 Q9PRT2
16	166.5	12.9	279	4 Q9IAF7
17	165	12.8	2225	5 Q9P881
18	161	439	12.5	4 Q9629
19	160.5	474	13	4 Q9IK94

ALIGNMENTS

RESULT	1	PRELIMINARY?	PRT:	331 AA.
ID	Q9UEK8			
AC	Q9UEK8;			
DT	01-MAY-2000	(TREMBLrel. 13; created)		
DT	01-JUN-2001	(TREMBLrel. 13; last sequence update)		
DE		EQUISTATIN PRECURSOR.		
GN				
BO1				
GO				
OC				
OC				
OC				
OX				
RN				
1]				
RP				
RA				
RT				
BIOL				
CH				
DR				
EMBL				
NCBI_TAXID				
INTERPRO				
DR				
PRION				
DR				
PROSITE				
PSD0484				
THYROGLOBULIN				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				
101				
102				
103				
104				
105				
106				
107				
108				
109				
110				
111				
112				
113				
114				
115				
116				
117				
118				
119				
120				
121				
122				
123				
124				
125				
126				
127				
128				
129				
130				
131				
132				
133				
134				
135				
136				
137				
138				
139				
140				
141				
142				
143				
144				
145				
146				
147				
148				
149				
150				
151				
152				
153				
154				
155				
156				
157				
158				
159				
160				
161				
162				
163				
164				
165				
166				
167				
168				
169				
170				
171				
172				
173				
174				
175				
176				
177				
178				
179				
180				
181				
182				
183				
184				
185				
186				
187				
188				
189				
190				
191				
192				
193				
194				
195				
196				
197				
198				
199				
200				
201				
202				
203				
204				
205				
206				
207				
208				
209				
210				
211				
212				
213				
214				
215				
216				
217				
218				
219				
220				
221				
222				
223				
224				
225				
226				
227				
228				
229				
230				
231				
232				
233				
234				
235				
236				
237				
238				
239				
240				
241				
242				
243				
244				
245				
246				
247				
248				
249				
250				
251				
252				
253				
254				
255				
256				
257				
258				
259				
260				
261				
262				
263				
264				
265				
266				
267				
268				
269				
270				
271				
272				
273				
274				
275				
276				
277				
278				
279				
280				
281				
282				
283				
284				
285				
286				
287				
288				
289				
290				
291				
292				
293				
294				
295				
296				
297				
298				
299				
300				
301				
302				
303				
304				
305				
306				
307				
308				
309				
310				
311				
312				
313				
314				
315				
316				
317				
318				
319				

RA Arneberg A.C.; Beas F.;
 RT "Studies on the structures of the normal and abnormal goat
 thyroglobulin genes";
 RT Biochimie 7:211-221(1989).
 RL EMBL; X14991; CAA33105; 1.
 DR InterPro; IPR000716; Thyroglobulin_1.
 DR Pfam; PF0006; thyroglobulin_1; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 2.
 FT NON-TER 212 AA; 23791 MW; 100035E08E8B9B8 CRC64;
 SQ SEQUENCE 212 AA; 23791 MW; 100035E08E8B9B8 CRC64;

Query Match 21.7%; Score 279; DB 6; Length 212;
 Best Local Similarity 39.2%; Pred No. 8; 9e-22;
 Matches 62; Conservative 18; Mismatches 60; Indels 18; Gaps 6;
 DR 4B YVPOCADGSFDTYOGKGDGASCWCVDAQREYPSRQRP---AACSFQLOQ 102

QY 112 I----WNPWCWGPSSKADSFEDYQCCASNGECYCVCVKKGKELETRORGRIT-CER 165
 DR 103 ILLSSYINSTASYLPOQSDGDYSPVQCDRLRRQCRQVPC 162

QY 166 HLSCBCE-EARAKHSNSRVERFVERCLDGDSNPYVC 202
 DR 163 --SCEFRNRLHVGVRSP--FQCSFDGAFRPVC 194

DR SMART; SM00211; TI; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 3.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 3.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.

FT SIGNAL 1 32 POTENTIAL.

SQ SEQUENCE 231 AA; 25410 MW; 07E256877D3356E9 CRC64;

Query Match 99.3%; Score 1278; DB 5; Length 231;
 Best Local Similarity 99.1%; Pred No. 5; 5.5e-128;
 Matches 229; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSONOAKFSKGIVVNITLFAITSTSEASITKCOLORASANSGLIGTVPOCKETG 60
 DR 1 MALSONOAKFSKGIVVNITLFAITSTSEASITKCOLORASANSGLIGTVPOCKETG 60

QY 61 EFEKQCGCSTSYCWVDEKGKIRSPDCKRKAATLCOMMUNITYVPGWCG 120
 DR 61 EFEKQCGCSTSYCWVDEKGKIRSPDCKRKAATLCOMMUNITYVPGWCG 120

QY 121 PPSKADGSEDFQOCASNGECYDKKGKELETRQGRGPCEHRLSCEARIKHSN 180
 DR 121 PPSKADGSEDFQOCASNGECYDKKGKELETRQGRGPCEHRLSCEARIKHSN 180

QY 181 SLRVEMFVPECFEDGSYNPVQCMSTYCWVDEGGVVKVPGSDFVRKRPTC 231
 DR 181 SLRVEMFVPECFEDGSYNPVQCMSTYCWVDEGGVVKVPGSDFVRKRPTC 231

RESULT 3

ID Q28330 PRELIMINARY; PRT; 212 AA.

AC Q28330; 01-NOV-1996 (TREMBREL; 01' Created)
 DT 01-NOV-1996 (TREMBREL; 01' Last sequence update)
 DT 01-JUN-2001 (TREMBREL; 17' Last annotation update)

DE DNA FOR THYROGLOBULIN (Tg) (FRAGMENT).

OS Capra hircus (Goat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Caprinae; Capra.

OC Bovidae; Caprinae; Capra.

DR [1] Sequence from N.A. MEDLINE=82020610; PubMed=2495821; RAX Van Ommeren G.J.B., de Vijlder J.J.M., Sterk A., Mercoken L.O.Y.,

RA Arneberg A.C.; Beas F.;
 RT "Studies on the structures of the normal and abnormal goat
 thyroglobulin genes";
 RT Biochimie 7:211-221(1989).
 RL EMBL; X14991; CAA33105; 1.
 DR InterPro; IPR000716; Thyroglobulin_1.
 DR Pfam; PF0006; thyroglobulin_1; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 3.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 3.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR SMART; SM00211; TI; 2.
 DR PROSITE; PS00018; PE_HAND; UNKNOWN_1.
 DR Calcium-binding; Signal; Signal; Signal.
 FT CHATN 22 446 SMOC-2.
 SQ SEQUENCE 446 AA; 49018 MW; 6C3303AE7212CC2A CRC64;

Query Match 20.2%; Score 259.5; DB 11; Length 446;
 Best Local Similarity 28.9%; Pred No. 2.3e-39;
 Matches 58; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

QY 41 QASANSLIGTVPOCKETGEFEKQCGCSTSYCWVDEGGVVKVPGSDFVRKRPTC 99

DR PROSITE; PS00018; EF HAND; UNKNOWN_2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
 KW Calcium-binding.
 SQ SEQUENCE 613 AA; 68955 MW; DE7EDC32B73F8A1A CRC64;

Query Match 13.4%; Score 173; DB 5; Length 613;
 Best Local Similarity 19.0%; Pred. No. 5.2e-10; Matches 70; Conservative 37; Mismatches 74; Indels 188; Gaps 14; QY 50 GTVYPPQCKETGEPEEKQWGSTGCWCVEDGEI----- 84
 DB 95 GFVFPVRKRDGNFAAMOCYGNNG-CWCSDSOSGRPIADDNKQFRRKKGKLRCRANRRRL 154
 QY 85 LGKIRGSPD-----CSR-----RKAALT-- 103
 DB 155 ASHQIGYKNPDTASKGSSEAGSTAHRTCSKSDRSQFNTNLMMERNEAQSSFFRQPSLDS 214
 QY 104 -----LCOMMQLAINV-PWCGGPSCK-----ADGSFDEV 133
 DB 215 HILEWQFSKLDTNGNKLDRQEIRELKVLRRNVKPRRCGRTFGKYCDVT DANLNWLE 274
 QY 134 OCCAS-----NECCYC 144
 DB 275 SVCFTEFHNSAVNLLASSATAAPPHSTHVSIHNNVNHGHRGHNTGTGVNPHSY 334
 QY 145 VDKKGKEJ-----EGTROGR--PTCERHL-----S 168
 DB 335 EDASGESEHDNYEDSATGGGSEEDDSQGADPSSRTHIPSILMLNSKPEASQDLENDS 394
 QY 169 EC---EERAKHNSLNRVEMYFPECLEDFGSYNPVCW---PSTGYWCVDG-GVXPGS 222
 DB 395 NCWMDQSTVLEOGHGGKSVLVPQCLFDGQRIOCYSSSTSVCWCVNEDTGKSIPGT 454
 QY 223 DVEFKRPC 231
 DB 455 SVKNKRPOC 463

RESULT 12

Q9VB21 PRELIMINARY; PRT; 1511 AA.

AC Q9VB21; DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG5639 PROTEIN.
 GN OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyoidea; Drosophilidae; Drosophila.
 RN [1] TaxID=7227;
 RN SEQUENCE FROM N.A.
 RC SPRAIN-BERKELEY;
 RX MEDLINE=20196005; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandt R.C., Rogers Y.H.C., Blazquez R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boisnakov S.,
 RA Botkova D., Botchan M.R., Bouck J., Broitstein P., Broitstein P.,
 RA Burts K.C., Busam D.A., Butler H., Cadine E., Center A., Chanda T.,
 RA Cherry J.M., Carley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Belcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrilian A.E., Garg N.S., Gelbart W.M., Glasser K.,

Query Match 13.3%; Score 171.5; DB 5; Length 151;
 Best Local Similarity 19.7%; Pred. No. 2e-09; Matches 72; Conservative 36; Mismatches 95; Indels 163; Gaps 17; QY 25 CAITSTEAL-TKQQQLA----SANGLIGHTY--VPQC-KETGEFEKEEKOWGSTGCYCW 75
 DB 823 CGTQCVDPOLIKTACQHQIAQLOHOSSELGIPATQAMAQCDPNNGKWNQVCG-SPQGHCW 881
 QY 76 CVDEDEKGKETLGTWTR-----GSPDCSRKAAL 102
 DB 882 CGDQGKILPGTRVKSPATPKCOENNSFACPKTNCSLECESYQMSNGCPCICRNICN 941
 QY 103 TL-----COMMAQAIINV-----GWGCP----- 121
 DB 942 EVSCSPHERECQLISVCVDPSPCPKMPICVPRASICPEGNPLQGDLDMSCGPHNHEVC 1001
 QY 122 --PSCKADGSEFEVQCASENGE-CY-----CVDKKGKELEGTR----- 156
 DB 1002 PTTHSQCNPVNURGVECSKTRDVCFSDMDGCLAGTKSENRSTRFSPANKCLPVVI 1061
 QY 157 QGGRPTCERH-----1SECEEARIKHNSLNRV-----MFVBC-LE 193
 DB 1062 DSAPACQTKNLFHNELACNSVCPVLTQCTERLKL--NNLAQRGHSSWFQPCRDPV 1118
 QY 194 DGSSYNPVQCW-----PSYGCMVDEGSYKVVGSDVR 225

Mon Mar 11 08:00:38 2002

us-09-445-480a-2.rspt

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:52 ; Search time 17.1 Seconds
 (without alignments)
 102.9.025 Million cell updates/sec

Title: US-09-445-480A-2
 Perfect score: 1287
 Sequence: 1 MALSQNQAKFSKGVVMIWV.....VDBGGVKVPGSDVRFKRPTC 231

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : PIR_68:
 1: pir1:
 2: pir2:
 3: pir3:
 4: pir4:
 5: pir5:
 6: pir6:
 7: pir7:
 8: pir8:
 9: pir9:
 10: pir10:
 11: pir11:
 12: pir12:
 13: pir13:
 14: pir14:
 15: pir15:
 16: pir16:
 17: pir17:
 18: pir18:
 19: pir19:
 20: pir20:
 21: pir21:
 22: pir22:
 23: pir23:
 24: pir24:
 25: pir25:
 26: pir26:
 27: pir27:
 28: pir28:
 29: pir29:
 30: pir30:
 31: pir31:
 32: pir32:
 33: pir33:
 34: pir34:
 35: pir35:
 36: pir36:
 37: pir37:
 38: pir38:
 39: pir39:
 40: pir40:
 41: pir41:
 42: pir42:
 43: pir43:
 44: pir44:
 45: pir45:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	295	22.9	2769	1 UIBO thyroglobulin precursor - bovine
2	279.5	21.7	2 A36124	thyroglobulin precursor - bovine
3	279	21.7	2 146176	thyroglobulin precursor - bovine
4	271	21.1	1 UIHU	thyroglobulin precursor - bovine
5	245.5	19.9	1 3736	thyroglobulin precursor - bovine
6	238.5	18.5	1 S31213	thyroglobulin precursor - bovine
7	165	12.8	2 2225	thyroglobulin precursor - bovine
8	159	12.4	2 S33293	thyroglobulin precursor - bovine
9	158.5	12.3	2 T32497	thyroglobulin precursor - bovine
10	154	12.0	1 NMMSND	thyroglobulin precursor - bovine
11	153	11.9	2 S04362	thyroglobulin precursor - bovine
12	153	11.9	1 MMHUND	thyroglobulin precursor - bovine
13	150	11.7	2 170687	thyroglobulin precursor - bovine
14	135	10.5	1 HLRHMSG	thyroglobulin precursor - bovine
15	132.5	10.3	2 A45403	thyroglobulin precursor - bovine
16	129.5	10.1	2 148599	thyroglobulin precursor - bovine
17	129	10.0	2 148603	thyroglobulin precursor - bovine
18	128.5	10.0	2 JC1464	thyroglobulin precursor - bovine
19	124.5	9.7	2 147031	thyroglobulin precursor - bovine
20	122.5	9.5	1 B37252	thyroglobulin precursor - bovine
21	122.5	9.5	2 JC4584	thyroglobulin precursor - bovine
22	122.5	9.5	2 A53748	thyroglobulin precursor - bovine
23	122.5	9.4	2 148604	thyroglobulin precursor - bovine
24	118.5	9.2	2 A25964	thyroglobulin precursor - bovine
25	118.5	9.2	2 JC1463	thyroglobulin precursor - bovine
26	114.5	8.9	2 T18355	thyroglobulin precursor - bovine
27	114	8.9	1 S23009	thyroglobulin precursor - bovine
28	114	8.8	1 MNNSB1	thyroglobulin precursor - bovine
29	113.5	8.8	2 A39842	thyroglobulin precursor - bovine

ALIGNMENTS

RESULT	1
R;Pampl, J.; Christophe, D.; Pohl, V.; Vassart, G.	R;Pampl, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987	J. Mol. Biol. 196, 769-779, 1987
A;Title: Structural organization of the 5' region of the thyroglobulin gene. Evi	A;Title: Structural organization of the 5' region of the thyroglobulin gene. Evi
A;Reference number: S03422; MUID:88062712	A;Reference number: S03422; MUID:88062712
A;Accession: S05566	A;Accession: S05566
A;Molecule type: DNA	A;Molecule type: DNA
A;Residues: 1002-1205, 'R', 1207-1209 <PAR>	A;Residues: 1002-1205, 'R', 1207-1209 <PAR>
A;Cross-references: EMBL:X06071; MUID:9778; PID:92204111	A;Cross-references: EMBL:X06071; MUID:9778; PID:92204111
R;Rawitch, A.B.; Litwer, M.R.; Gregg, J.; Turner, C.D.; Rouse, J.B.; Hamilton, J	R;Rawitch, A.B.; Litwer, M.R.; Gregg, J.; Turner, C.D.; Rouse, J.B.; Hamilton, J
Biochem. Biophys. Res. Commun. 118, 423-429, 1984	Biochem. Biophys. Res. Commun. 118, 423-429, 1984
A;Title: The isolation of identical thyroxine containing amino acid sequences from bovine thyroglobulin. A;Reference number: A32735; MUID:84153804	A;Title: The isolation of identical thyroxine containing amino acid sequences from bovine thyroglobulin. A;Reference number: A32735; MUID:84153804
A;Accession: A32735	A;Accession: A32735
A;Molecule type: protein	A;Molecule type: protein
A;Residues: 20-38 <RAW>	A;Residues: 20-38 <RAW>
R;Gentile, F.; Salvatore, G.	R;Gentile, F.; Salvatore, G.
Bur. J. Biochem. 218, 603-621, 1993	Bur. J. Biochem. 218, 603-621, 1993
A;Title: Preferential sites of proteolytic cleavage of bovine, human and rat thy	A;Title: Preferential sites of proteolytic cleavage of bovine, human and rat thy
A;Reference number: S39431; MUID:94094855	A;Reference number: S39431; MUID:94094855
A;Accession: S39431	A;Accession: S39431
A;Molecule type: protein	A;Molecule type: protein
A;Residues: 20-27, 259-265, 521-526, 'X', 528, 539-560, 1161-1168, 'XX', 1171-1172, 1203-1204, 1205-1206, 1207-1208, 1209-1210, 1211-1212, 1213-1214, 1215-1216, 1217-1218, 1219-1220, 1221-1222, 1223-1224, 1225-1226, 1227-1228, 1229-1230, 1231-1232, 1233-1234, 1235-1236, 1237-1238, 1239-1240, 1241-1242, 1243-1244, 1245-1246, 1247-1248, 1249-1250, 1251-1252, 1253-1254, 1255-1256, 1257-1258, 1259-1260, 1261-1262, 1263-1264, 1265-1266, 1267-1268, 1269-1270, 1271-1272, 1273-1274, 1275-1276, 1277-1278, 1279-1280, 1281-1282, 1283-1284, 1285-1286, 1287-1288, 1289-1290, 1291-1292, 1293-1294, 1295-1296, 1297-1298, 1299-1299, 1300-1301, 1301-1302, 1302-1303, 1303-1304, 1304-1305, 1305-1306, 1306-1307, 1307-1308, 1308-1309, 1309-1310, 1310-1311, 1311-1312, 1312-1313, 1313-1314, 1314-1315, 1315-1316, 1316-1317, 1317-1318, 1318-1319, 1319-1320, 1320-1321, 1321-1322, 1322-1323, 1323-1324, 1324-1325, 1325-1326, 1326-1327, 1327-1328, 1328-1329, 1329-1330, 1330-1331, 1331-1332, 1332-1333, 1333-1334, 1334-1335, 1335-1336, 1336-1337, 1337-1338, 1338-1339, 1339-1340, 1340-1341, 1341-1342, 1342-1343, 1343-1344, 1344-1345, 1345-1346, 1346-1347, 1347-1348, 1348-1349, 1349-1350, 1350-1351, 1351-1352, 1352-1353, 1353-1354, 1354-1355, 1355-1356, 1356-1357, 1357-1358, 1358-1359, 1359-1360, 1360-1361, 1361-1362, 1362-1363, 1363-1364, 1364-1365, 1365-1366, 1366-1367, 1367-1368, 1368-1369, 1369-1370, 1370-1371, 1371-1372, 1372-1373, 1373-1374, 1374-1375, 1375-1376, 1376-1377, 1377-1378, 1378-1379, 1379-1380, 1380-1381, 1381-1382, 1382-1383, 1383-1384, 1384-1385, 1385-1386, 1386-1387, 1387-1388, 1388-1389, 1389-1390, 1390-1391, 1391-1392, 1392-1393, 1393-1394, 1394-1395, 1395-1396, 1396-1397, 1397-1398, 1398-1399, 1399-1400, 1400-1401, 1401-1402, 1402-1403, 1403-1404, 1404-1405, 1405-1406, 1406-1407, 1407-1408, 1408-1409, 1409-1410, 1410-1411, 1411-1412, 1412-1413, 1413-1414, 1414-1415, 1415-1416, 1416-1417, 1417-1418, 1418-1419, 1419-1420, 1420-1421, 1421-1422, 1422-1423, 1423-1424, 1424-1425, 1425-1426, 1426-1427, 1427-1428, 1428-1429, 1429-1430, 1430-1431, 1431-1432, 1432-1433, 1433-1434, 1434-1435, 1435-1436, 1436-1437, 1437-1438, 1438-1439, 1439-1440, 1440-1441, 1441-1442, 1442-1443, 1443-1444, 1444-1445, 1445-1446, 1446-1447, 1447-1448, 1448-1449, 1449-1450, 1450-1451, 1451-1452, 1452-1453, 1453-1454, 1454-1455, 1455-1456, 1456-1457, 1457-1458, 1458-1459, 1459-1460, 1460-1461, 1461-1462, 1462-1463, 1463-1464, 1464-1465, 1465-1466, 1466-1467, 1467-1468, 1468-1469, 1469-1470, 1470-1471, 1471-1472, 1472-1473, 1473-1474, 1474-1475, 1475-1476, 1476-1477, 1477-1478, 1478-1479, 1479-1480, 1480-1481, 1481-1482, 1482-1483, 1483-1484, 1484-1485, 1485-1486, 1486-1487, 1487-1488, 1488-1489, 1489-1490, 1490-1491, 1491-1492, 1492-1493, 1493-1494, 1494-1495, 1495-1496, 1496-1497, 1497-1498, 1498-1499, 1499-1500, 1500-1501, 1501-1502, 1502-1503, 1503-1504, 1504-1505, 1505-1506, 1506-1507, 1507-1508, 1508-1509, 1509-1510, 1510-1511, 1511-1512, 1512-1513, 1513-1514, 1514-1515, 1515-1516, 1516-1517, 1517-1518, 1518-1519, 1519-1520, 1520-1521, 1521-1522, 1522-1523, 1523-1524, 1524-1525, 1525-1526, 1526-1527, 1527-1528, 1528-1529, 1529-1530, 1530-1531, 1531-1532, 1532-1533, 1533-1534, 1534-1535, 1535-1536, 1536-1537, 1537-1538, 1538-1539, 1539-1540, 1540-1541, 1541-1542, 1542-1543, 1543-1544, 1544-1545, 1545-1546, 1546-1547, 1547-1548, 1548-1549, 1549-1550, 1550-1551, 1551-1552, 1552-1553, 1553-1554, 1554-1555, 1555-1556, 1556-1557, 1557-1558, 1558-1559, 1559-1560, 1560-1561, 1561-1562, 1562-1563, 1563-1564, 1564-1565, 1565-1566, 1566-1567, 1567-1568, 1568-1569, 1569-1570, 1570-1571, 1571-1572, 1572-1573, 1573-1574, 1574-1575, 1575-1576, 1576-1577, 1577-1578, 1578-1579, 1579-1580, 1580-1581, 1581-1582, 1582-1583, 1583-1584, 1584-1585, 1585-1586, 1586-1587, 1587-1588, 1588-1589, 1589-1590, 1590-1591, 1591-1592, 1592-1593, 1593-1594, 1594-1595, 1595-1596, 1596-1597, 1597-1598, 1598-1599, 1599-1599, 1599-1600, 1600-1601, 1601-1602, 1602-1603, 1603-1604, 1604-1605, 1605-1606, 1606-1607, 1607-1608, 1608-1609, 1609-1610, 1610-1611, 1611-1612, 1612-1613, 1613-1614, 1614-1615, 1615-1616, 1616-1617, 1617-1618, 1618-1619, 1619-1620, 1620-1621, 1621-1622, 1622-1623, 1623-1624, 1624-1625, 1625-1626, 1626-1627, 1627-1628, 1628-1629, 1629-1630, 1630-1631, 1631-1632, 1632-1633, 1633-1634, 1634-1635, 1635-1636, 1636-1637, 1637-1638, 1638-1639, 1639-1640, 1640-1641, 1641-1642, 1642-1643, 1643-1644, 1644-1645, 1645-1646, 1646-1647, 1647-1648, 1648-1649, 1649-1650, 1650-1651, 1651-1652, 1652-1653, 1653-1654, 1654-1655, 1655-1656, 1656-1657, 1657-1658, 1658-1659, 1659-1660, 1660-1661, 1661-1662, 1662-1663, 1663-1664, 1664-1665, 1665-1666, 1666-1667, 1667-1668, 1668-1669, 1669-1670, 1670-1671, 1671-1672, 1672-1673, 1673-1674, 1674-1675, 1675-1676, 1676-1677, 1677-1678, 1678-1679, 1679-1680, 1680-1681, 1681-1682, 1682-1683, 1683-1684, 1684-1685, 1685-1686, 1686-1687, 1687-1688, 1688-1689, 1689-1690, 1690-1691, 1691-1692, 1692-1693, 1693-1694, 1694-1695, 1695-1696, 1696-1697, 1697-1698, 1698-1699, 1699-1699, 1699-1700, 1700-1701, 1701-1702, 1702-1703, 1703-1704, 1704-1705, 1705-1706, 1706-1707, 1707-1708, 1708-1709, 1709-1710, 1710-1711, 1711-1712, 1712-1713, 1713-1714, 1714-1715, 1715-1716, 1716-1717, 1717-1718, 1718-1719, 1719-1720, 1720-1721, 1721-1722, 1722-1723, 1723-1724, 1724-1725, 1725-1726, 1726-1727, 1727-1728, 1728-1729, 1729-1730, 1730-1731, 1731-1732, 1732-1733, 1733-1734, 1734-1735, 1735-1736, 1736-1737, 1737-1738, 1738-1739, 1739-1740, 1740-1741, 1741-1742, 1742-1743, 1743-1744, 1744-1745, 1745-1746, 1746-1747, 1747-1748, 1748-1749, 1749-1750, 1750-1751, 1751-1752, 1752-1753, 1753-1754, 1754-1755, 1755-1756, 1756-1757, 1757-1758, 1758-1759, 1759-1760, 1760-1761, 1761-1762, 1762-1763, 1763-1764, 1764-1765, 1765-1766, 1766-1767, 1767-1768, 1768-1769, 1769-1770, 1770-1771, 1771-1772, 1772-1773, 1773-1774, 1774-1775, 1775-1776, 1776-1777, 1777-1778, 1778-1779, 1779-1780, 1780-1781, 1781-1782, 1782-1783, 1783-1784, 1784-1785, 1785-1786, 1786-1787, 1787-1788, 1788-1789, 1789-1790, 1790-1791, 1791-1792, 1792-1793, 1793-1794, 1794-1795, 1795-1796, 1796-1797, 1797-1798, 1798-1799, 1799-1799, 1799-1800, 1800-1801, 1801-1802, 1802-1803, 1803-1804, 1804-1805, 1805-1806, 1806-1807, 1807-1808, 1808-1809, 1809-1810, 1810-1811, 1811-1812, 1812-1813, 1813-1814, 1814-1815, 1815-1816, 1816-1817, 1817-1818, 1818-1819, 1819-1820, 1820-1821, 1821-1822, 1822-1823, 1823-1824, 1824-1825, 1825-1826, 1826-1827, 1827-1828, 1828-1829, 1829-1830, 1830-1831, 1831-1832, 1832-1833, 1833-1834, 1834-1835, 1835-1836, 1836-1837, 1837-1838, 1838-1839, 1839-1840, 1840-1841, 1841-1842, 1842-1843, 1843-1844, 1844-1845, 1845-1846, 1846-1847, 1847-1848, 1848-1849, 1849-1850, 1850-1851, 1851-1852, 1852-1853, 1853-1854, 1854-1855, 1855-1856, 1856-1857, 1857-1858, 1858-1859, 1859-1860, 1860-1861, 1861-1862, 1862-1863, 1863-1864, 1864-1865, 1865-1866, 1866-1867, 1867-1868, 1868-1869, 1869-1870, 1870-1871, 1871-1872, 1872-1873, 1873-1874, 1874-1875, 1875-1876, 1876-1877, 1877-1878, 1878-1879, 1879-1880, 1880-1881, 1881-1882, 1882-1883, 1883-1884, 1884-1885, 1885-1886, 1886-1887, 1887-1888, 1888-1889, 1889-1890, 1890-1891, 1891-1892, 1892-1893, 1893-1894, 1894-1895, 1895-1896, 1896-1897, 1897-1898, 1898-1899, 1899-1899, 1899-1900, 1900-1901, 1901-1902, 1902-1903, 1903-1904, 1904-1905, 1905-1906, 1906-1907, 1907-1908, 1908-1909, 1909-1910, 1910-1911, 1911-1912, 1912-1913, 1913-1914, 1914-1915, 1915-1916, 1916-1917, 1917-1918, 1918-1919, 1919-1920, 1920-1921, 1921-1922, 1922-1923, 1923-1924, 1924-1925, 1925-1926, 1926-1927, 1927-1928, 1928-1929, 1929-1930, 1930-1931, 1931-1932, 1932-1933, 1933-1934, 1934-1935, 1935-1936, 1936-1937, 1937-1938, 1938-1939, 1939-1940, 1940-1941, 1941-1942, 1942-1943, 1943-1944, 1944-1945, 1945-1946, 1946-1947, 1947-1948, 1948-1949, 1949-1950, 1950-1951, 1951-1952, 1952-1953, 1953-1954, 1954-1955, 1955-1956, 1956-1957, 1957-1958, 1958-1959, 1959-1960, 1960-1961, 1961-1962, 1962-1963, 1963-1964, 1964-1965, 1965-1966, 1966-1967, 1967-1968, 1968-1969, 1969-1970, 1970-1971, 1971-1972, 1972-1973, 1973-1974, 1974-1975, 1975-1976, 1976-1977, 1977-1978, 1978-1979, 1979-1980, 1980-1981, 1981-1982, 1982-1983, 1983-1984, 1984-1985, 1985-1986, 1986-1987, 1987-1988, 1988-1989, 1989-1990, 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1994-1995, 1995-1996, 1996-1997, 1997-1998, 1998-1999, 1999-1999, 1999-2000, 2000-2001, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021, 2021-2022, 2022-2023, 2023-2024, 2024-2025, 2025-2026, 2026-2027, 2027-2028, 2028-2029, 2029-2030, 2030-2031, 2031-2032, 2032-2033, 2033-2034, 2034-2035, 2035-2036, 2036-2037,	

R; Rawitch, A.B.; Pollock, H.G.; Yang, S.X.
 Arch. Biochem. Biophys. 300, 271-279, 1993
 A; Title: Thyroglobulin glycosylation: location and nature of the N-linked oligosaccharid
 A; Accession: S29734
 A; Molecule type: protein
 A; Residues: 106-109, 'X', 111-117, 475-477, 'N', 479-482, 'X', 484-485, 492-494, 'X', 496-502, 839-
 252-254, 2272-2281 <RA3>
 R; Ricketts, M.H.; Simons, M.
 Proc. Natl. Acad. Sci. U.S.A. 84, 3181-3184, 1987
 A; Title: A nonsense mutation causes hereditary goitre in the Afrikander cattle and umas
 A; Reference number: 145981; MUID:87204101
 A; Status: preliminary; translated from GB/EMBL/DDJB
 A; Molecule type: DNA
 A; Accession: 145981
 A; Status: preliminary; translated from GB/EMBL/DDJB
 A; Cross-references: GB:MI6448; NID:9163743; PIDN:AAA30777_1; PID:9552340
 C; Comment: The thyroglobulin molecule is a dimeric glycoprotein specific to the thyroid
 C; Genetics:
 A; Introns: 1047/1; 1073/1; 1110/3; 1145/1
 A; Note: the list of introns may be incomplete
 C; Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homologous domain
 C; Keywords: duplication; glycoprotein; homodimer; iodine; thyroid gland; thyroid hormone
 F; 1-19/Domain: signal sequence #status predicted
 F; 20-2769/Product: thyroglobulin #status predicted <TGL>
 F; 34-92/Domain: thyroglobulin type I repeat homology <TH01>
 F; 96-160/Domain: thyroglobulin type I repeat homology #status atypical <TH02>
 F; 164-297/Domain: thyroglobulin type I repeat homology #status atypical <TH03>
 F; 301-358/Domain: thyroglobulin type I repeat homology <TH04>
 F; 607-657/Domain: thyroglobulin type I repeat homology <TH05>
 F; 661-725/Domain: thyroglobulin type I repeat homology <TH06>
 F; 729-921/Domain: thyroglobulin type I repeat homology #status atypical <TH07>
 F; 923-1073/Domain: thyroglobulin type I repeat homology #status atypical <TH08>
 F; 1077-1145/Domain: thyroglobulin type I repeat homology <TH09>
 F; 1149-1210/Domain: thyroglobulin type I repeat homology <TH10>
 F; 158-1471/Region: type II repeat
 F; 1472-1488/Region: type II repeat
 F; 1605-1725/Region: type III repeat
 F; 1726-1893/Region: type III repeat
 F; 1894-1996/Region: type III repeat
 F; 1997-2130/Region: type III repeat
 F; 2131-2188/Region: type III repeat
 F; 2229-2727/Domain: cholinesterase homology <ACE>
 F; 24/Modified site: thyroxine (Tyr) #status experimental
 F; 110,483,495,853,947,1140,1365,1776,1870,2014,2123,2251,2296/Binding site: carbohydrate
 F; 198-Binding site: carbohydrate (Asn) (covalent) #status absent
 F; 2574/Modified site: thyroxine (Tyr) #status predicted
 F; 2588/Modified site: thyroxine (Tyr) #status predicted
 F; 2677/Modified site: triiodothyronine (Tyr) #status predicted

Query Match 22.9%; Score 295; DB 1; Length 2769;
 Best Local Similarity 33.5%; Pred. No. 3.1e-16;
 Matches 59; Conservative 21; Mismatches 64; Indels 52; Gaps 8;

QY 52 YVPOCKETGEEFEKQCMWGSTGYCWCVDDEKGKTRGSDCSRKKALTLQMMQAI 111
 Db 48 YVPQCAEDSFQTFQVCGKGASCMVQDADGREGVPGSQGPGRP----AACLSFCQLOKQQ 102

QY 112 T----VNTPGCWPSPSKAKGSDPEWVOCASNGECYCVDKKGKELEGTRQGRPT-CER 165
 Db 103 ILSSYVINSTATSYLPOCGSDGYSQVPCDRLRRQCMWCVDAAEGMEVYGTROQGRPARCP 162

QY 166 HLSCECERAKTAKAISNSLNLVEMFVPECLGEGDSYNPVOC----- 202
 Db 163 ---SCEIRNRRLIGVGDSR---PQCGPDGARPVQCKLWNITDMMFIDLVHSYSRFD 216

QY 203 -----WPS-TGICWCVPBGG 216
 Db 217 AFVTFFSSRSRFPEVSGCYCABDSQG 242

RESULT 2
 A36124
 thyroglobulin 2 precursor - rat
 C; Species: Rattus norvegicus (norway rat)
 C; Date: 14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change 11-Apr-1997
 C; Accession: A36124
 R; Graves, P.N.; Davies, T.F.
 Mol. Endocrinol. 4, 155-161, 1990
 A; Title: A second thyroglobulin messenger RNA species (rtg-2) in rat thyrocytes.
 A; Reference number: A36124; MUID:90220642
 A; Accession: A36124
 A; Status: preliminary
 A; Molecule type: mRNA
 A; Residues: 1-243 <GRA>
 A; Cross-references: GB:MI5965
 C; Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homologous domain
 F; 91-161/Domain: thyroglobulin type I repeat homology <THY1>
 F; 25/Modified site: thyroxine (Tyr) #status predicted

Query Match 21.7%; Score 279; DB 2; Length 212;
 Best Local Similarity 39.2%; Pred. No. 6.6e-16;
 Matches 62; Conservative 18; Mismatches 60; Indels 18; Gaps 6;

QY 52 YVPOCKETGEEFEKQCMWGSTGYCWCVDDEKGKTRGSDCSRKKALTLQMMQAI 111
 Db 48 YVPQCAEDSFQTFQVCGKGASCMVQDADGREGVPGSQGPGRP----AACLSFCQLOKQQ 102

F;110 496 1868 2121/Binding site: carbohydrate (Asn) (covalent) #status absent
 F;149/Modified site: dehydroalanine (Tyr) #status predicted
 F;2765/Modified site: triiodothyronine (Tyr) #status predicted

RESULT 6
 S31213
 nidoigen precursor - sea squirt (*Halocynthia roretzii*)
 nidoigen precursor - sea squirt (*Halocynthia roretzii*)
 nidoigen precursor - sea squirt (*Halocynthia roretzii*)

N

A

I

C

S

E

R

T

C

P

O

D

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

Query Match 12.8%; Score 165; DB 2; Length 2225;
 Best Local Similarity 26.4%; Pred. No. 1.2e-05; Matches 58; Conservative 28; Mismatches 68; Indels 66; Gaps 10;

Qy 29 STASLTIKQQLQASANGLT-GTVVPPQCGTEEEKCWGSITGWCYVMDENGKEITGT 87
 Db 380 TTRAPRRECERRASASSSSTRGGVPPACTASGFERNQRCVNTQIGLEPPNS 439

Qy 88 KIRGS-PDCSERKAKMLTCQMQAITAVNPFGWMGCPPSKADSF-----DEVQ--- 135
 Db 440 RTDGTRPDCSIQSSTIV-NIKECVGVSIFPGC-----HGSFQRYFINEDSQCEQF 491

Qy 136 ---CANGECCYVDKKKGELEGTRQGRERH-----SECEARIRAKHSNLRV 184

Db 492 TWSGGGGNNYY-----ESRACEDRCAAPPVGRLRCE----- 524

RESULT 8

Query Match 12.3%; Score 158.5; DB 2; Length 2150;
 Best Local Similarity 20.7%; Pred. No. 3.9e-05; Matches 67; Conservative 20.7%; Mismatches 100; Indels 121; Gaps 15;

N;Alternate names: glycosaminoglycan-bearing polypeptide
 C;Species: Homo sapiens (man)
 C;Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
 C;Accession: S33293; S27165; P2165; P3556; P3557; P3558; P3559; P3560; P3561
 R;Alliel, P.M.; Perin, J.P.; Jolles, P.; Bonnet, F.J.
 Eur. J. Biochem. 214, 347-350, 1993
 A;Title: Characterization of a multi-domain testicular proteoglycan resembling modulators of cell s
 A;Reference number: S33293; MUID:93285162
 A;Accession: S33293
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: I-439 <ALL>
 R;Bonnet, F.; Perin, J.P.; Maillet, P.; Jolles, P.; Alliel, P.M.
 Biochem. J. 288, 565-569, 1992
 A;Title: Characterization of a human seminal plasma glycosaminoglycan-bearing polypeptid
 A;Reference number: S27166; MUID:930998810
 A;Accession: S27166
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 370-382, 'X', 384-387, 'X', 389-392 <BON>
 C;Superfamily: Kazal proteinase inhibitor homology; thyroglobulin type I repeat homology
 F;132-180/Domain: Kazal proteinase inhibitor homology <KP16>
 F;313-376/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 12.4%; Score 159; DB 2; Length 439;
 Best Local Similarity 23.8%; Pred. No. 8.7e-06; Matches 40; Conservative 28; Mismatches 48; Indels 52; Gaps 6;

Qy 43 SANSGLIGTYVPOCKEKGEEFEKRCWGSTGCWCWVDEGSKELIGTKIRGSPDGSRRKAAL 102
 Db 324 SKGRSLIGAFPRCNGGYKATQCHGTSQGWCYDQKGMLAGSRKGAVSSEQ--- 380

Qy 103 TLCOMMATTIVNPWGCGPPSKCADSSFDDEVOCAGCASNCGECYCVDKKGLELEGTRQGRPT 162
 Db 381 -----ETSDF-----GSGGSVLLD---DIEVERELG--- 405

Qy 163 CERHLSECEEARIKAHNSNLRVEMFPECLLEDGSYNPVQCPWPSIGCY 210

Db 406 ---PKOKEGKLVRHRAVTD---DEDDEDDKDEEV---GYIN 439

RESULT 9

C;Accession: T32497
 R;Gesel, C.; Stelljes, L.
 Submitted to the EMBL Data Library, December 1997
 A;Description: The sequence of *C. elegans* cosmid C08G9.
 A;Reference number: Z21179
 A;Accession: T32497
 A;Status: preliminary; translated from GB/EMBL/DBBJ
 A;Molecule type: DNA
 A;Residues: 1-2150 <GET>
 A;Cross-references: EMBL:Ar036687; PIDN:AAB88311.1; GSPDB:GN00022; CESP:C08G9.2
 A;Experimental source: strain Bristol N2; clone c08g9
 C;Genetics:
 A;Gene: CESP:C08G9.2
 A;Map position: 4
 A;Introns: 242/1; 306/1; 340/3; 485/1; 545/1; 736/2; 791/1; 829/1; 886/3; 942/1;

Query Match 12.3%; Score 158.5; DB 2; Length 2150;
 Best Local Similarity 20.7%; Pred. No. 3.9e-05; Matches 67; Conservative 20.7%; Mismatches 100; Indels 121; Gaps 15;

Qy 22 FIACAITS-----TEASLTKKOOLASANSANGLIGTYVP--OCETGEEEEKCWG 69
 Db 1380 FLKQVQASGTMCSAPIRATACIHLIAFESNIASETIVNLIPFVQCTPEGFLRKQCD 1439

Qy 70 STGVCWCVDE-DEKEIIGTKI---RGSPCDSRRKAALTC----- 105
 Db 1440 RLNOQCVKDATGVEILGTRIFTIGQSPDCNLPKSCPFCISRCPCPYGIRTGSSCPAN 1499

Qy 106 -----QMMQALIVNVGWCGB---PSCADGSDDEVOC----- 135
 Db 1500 GVCECINCSTFNCPQNMELLARRVVCSTSNCPPDVPPFCV-----EVOCPIORDLYRN 1553

Qy 136 -----CASNEC-----YCVDKKCE----- 160
 Db 1554 ILCENDGGGKNSKCNVKLPKTEMGICHOIKTQVSSLVNLISPAIVAAPEVIEATTVDR 1613

Qy 161 ---PTCERHL---SECEARIK---AHSNLSRLEMFPECLLEDGSYNPVQCPWPSIGCY 209
 Db 1614 GNNSPVVWEAVPHSLTNCTTMIALYLHQNGAVLKSPLVCSRNQNYDQTC--DNKRC 1671

Qy 210 WCVDIE-GGKVVKPGPSDVRKRPTC 231
 Db 1672 WCVDLSELSEBIEHGTRKKTKNAC 1694

RESULT 10

MMSND
 nidoen precursor - mouse
 N;Alternate names: entactin
 C;Species: Mus musculus (house mouse)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Oct-2000
 C;Accession: S02730; A38788; PNU669; S03037; A38898; S14878; S14835; A42785; A571
 R;Durkin, M.E.; Chakravarti, S.; Bartos, B.B.; Liu, S.H.; Friedman, R.L.; Chung, J.; Cell Biol. 107, 2749-2756, 1988
 A;Title: Amino acid sequence and domain structure of entactin. Homology with epid
 A;Reference number: S02730; MUID:89079780
 A;Accession: S02730
 A;Molecule type: mRNA
 A;Residues: 1-1245 <DUR>
 A;Cross-references: EMBL:X14194; NID:950838; PIDN:CAA32408.1; PID:950839
 A;Accession: A38788
 A;Molecule type: protein
 A;Residues: 29-30, 'X', 32-40 <DU2>
 R;Durkin, M.E.; Liu, S.H.; Reing, J.; Chung, A.E.
 Gene 132, 261-266, 1993
 A;Title: Characterization of the 5' end of the mouse Ent gene encoding the baseme
 A;Reference number: PNU669; MUID:94040771
 A;Accession: PNU669
 A;Molecule type: DNA
 A;Cross-references: GB:L117322; GB:L117323; GB:L117324; NID:9410235; PIDN:AAA77652.1
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 T32497 hypothetical protein C08G9.2 - *Caenorhabditis elegans*
 T32497 hypothetical protein C08G9.2 - *Caenorhabditis elegans*
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

EMBO J. 8, 65-72, 1989
 A;Title: Amino acid sequence of mouse nidogen, a multidomain basement membrane protein w
 A;Reference number: S03637; MUID:89231638
 A;Accession: S03637
 A;Molecule type: mRNA
 A;Residues: 1-169, 'L'
 A;Cross-references: EMBL:X14480; NID:953383; PTDN:CAA32642.1; PID:953384
 A;Accession: A3879
 A;Molecule type: protein
 A;Residues: 29-53;130-167,'G',169-170,'V',172-173,'A',175;210-223;225-240;252-295;318-33
 A;Residues: 55-761;773-796;809-836;879-902;955-970;959-969;976-990;993;1003-1098;1100-1122;1131-1141;114
 R;Mann, K.; Deutzmman, R.; Timpl, R.
 Eur. J. Biochem. 178, 71-80, 1988
 A;Title: Characterization of proteolytic fragments of the laminin-nidogen complex and th
 A;Reference number: S08095; MUID:89078415
 A;Accession: S14878
 A;Molecule type: protein
 A;Residues: 29-49;299-336,'F',338-649-694,'X';809-836;1003-1165;1066-093 <MA2>
 R;Paulsson, M.; Deutzmman, R.; Dziadek, M.; Nowack, H.; Timpl, R.; Weber, S.; Engel, J.
 Eur. J. Biochem. 155, 467-478, 1986
 A;Title: Purification and structural characterization of intact and fragmented nidogen
 A;Reference number: S14836; MUID:8619247
 A;Accession: S14836
 A;Molecule type: protein
 A;Residues: 29-38;222-240;326-330,'A',332-335;379-387,'N';649-658 <PAU>
 A;Note: 338-Gly was also found
 R;Aeschlimann, D.; Paulsson, M.; Mann, K.
 J. Biol. Chem. 267, 11316-1121, 1992
 A;Title: Identification of Glu(726) in nidogen as the amine acceptor in transglutaminase
 A;Reference number: A42785; MUID:9228840
 A;Accession: A42785
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 350-369;660-671;746-772 <AES>
 R;Durkin, M.E.; Rewar, U.M.; Chung, A.B.
 Genomics 26, 219-228, 1995
 A;Title: Exon organization of the mouse enectatin gene corresponds to the structural doma
 A;Reference number: A57168; MUID:95324912
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 252-966,'A',968-1205,'Q',1206-1245 <DU4>
 A;Cross-references: GB:X83041; GB:X83042; GB:X83043; GB:X83078; GB:X83079; GB:X83080; GB
 X83091; GB:X83092; GB:X83093
 C;Comment: This protein is a basement membrane glycoprotein that forms a complex with la
 C;Genetics:
 A;Gene: ENT; Nid
 A;Map Position: 13
 C;Superfamily: nidogen; EGF homology; LDL receptor YWTD-containing repeat homology; thy
 C;Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding; cc
 F;1-28/Domain: signal sequence #status predicted <SIG>
 F;29-1245/Product: nidogen #status experimental <MAT>
 A;Accession: A57168
 F;43-54/Domain: calcium binding #status predicted <CA1>
 F;278-289/Domain: calcium binding #status predicted <CA2>
 F;388-423/Domain: EGF homology <EG1>
 F;670-706/Domain: EGF homology <EG2>
 F;701-702/Region: cell attachment (R-G-D) motif
 F;712-748/Domain: EGF homology <EG3>
 F;760-798/Domain: EGF homology <EG4>
 F;804-837/Domain: EGF homology <EG5>
 F;847-917/Domain: thyroglobulin type I repeat homology <THY1>
 F;988-1030/Domain: LDL receptor YWTD-containing repeat homology <YW1>
 F;1031-1073/Domain: LDL receptor YWTD-containing repeat homology <YW2>
 F;1074-1118/Domain: LDL receptor YWTD-containing repeat homology <YW3>
 F;1116-1158/Domain: LDL receptor YWTD-containing repeat homology <YW4>
 F;1159-1195/Domain: LDL receptor YWTD-containing repeat homology <YW5>
 F;1210-1241/Domain: EGF homology <EG6>
 F;187/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;290-295/Binding site: sulfate (Tyr) (covalent) #status predicted
 F;47-51/Modified site: carbohydrate (Asn) (covalent) #status experimental
 F;727-817/Modified site: erythro-beta hydroxyasparagine (Asn) #status predicted
 F;754/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino (laminin)) #status exper

Query Match 12.0%; Score 154; DB 1; Length 1245;
 Best Local Similarity 27.8%; Pred. No. 5.6e-05; Gaps 10;
 Matches 58; Conservative 27; Mismatches 84; Indels 40; Gaps 10;

Qy	46	SGLIGTYVPOQKETGFEFKCQMGSTGCVWCGVYDKEILOTKRKGSPCSRKALPLIC	105
Db	726	NLILPGTFRCCEVCVEGYHFSDR--GT-----CVAEDORPINYCETGLHNCD-----IP	770
Qy	105	QMKQALIYN-----VPWCGPSPCKADGSEDEV--CCASNEY-----CVDKK 148	
Db	771	QRAQCTYNGGSYTCSCLPQFGSG--DGRACRVDDECOHSRCHPDARCYNPGSFTCOCKP	828
Qy	149	GKELEGTR--QQGRPTCRHSLCEEARIKAHNSLRVE-MFPECLEDGSNPVQWCW 203	
Db	829	GYGDGFKCMGPEVSKTRCQLERHLHGAGGDAOPTLQMFPCQCBYGHVPTOCH 888	
Qy	204	PSGTYCMVQDVDEGGVKVPGSDV-RFKRPTC 231	
Db	889	HSTGYCWCVDRDGRELEGSRTPGMRPPC 917	

RESULT 11
 S04362
 class II histocompatibility antigen-associated gamma chain long splice form - rat
 N; Contains: class II histocompatibility antigen-associated gamma chain, short splice
 C; Species: Rattus norvegicus (Norway rat)
 C; Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 23-Jul-1999
 C; Accession: S04362; S02162
 R; McKnight, A.J.; Mason, D.W.; Barclay, A.N.
 Nucleic Acids Res. 17, 3983-3984, 1989
 A; Title: Sequence of a rat MHC class II-associated invariant chain cDNA clone contains
 A; Reference number: S04362; MVID:89282409
 A; Accession: S04362
 A; Molecule type: mRNA
 A; Residues: 1-280 <MGR>
 A; Cross-references: EMBL:X10304; NID:9566497; PIDN:CAA31450.1; PID:9566498
 R; Henkels, W.; Synts, J.; Reske, K.
 Nucleic Acids Res. 16, 1822, 1988
 A; Title: Nucleotide sequence of rat invariant gamma chain cDNA clone pLR-gamma-34.
 A; Reference number: S02182; MVID:89098337
 A; Accession: S02182
 A; Molecule type: mRNA
 A; Residues: 1-192,257-280 <HEN>
 A; Cross-references: EMBL:X10304; NID:9566497; PIDN:CAA31450.1; PID:9566498
 R; Henkels, W.; Synts, J.; Reske, K.
 Nucleic Acids Res. 16, 1822, 1988
 A; Note: the authors translated the codon ATC for residue 16 as Asn, CTG for residue
 C; Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin
 C; Keywords: alternative splicing; transmembrane protein
 F; 197-255/Domain: thyroglobulin type I repeat homology <THY1>
 Query Match 11.9%; Score 153; DB 2; Length 280;
 Best Local Similarity 38.4%; Pred. No. 1.8e-05; Gaps 1;
 Matches 33; Conservative 10; Mismatches 35; Indels 8; Gaps 1;

Qy	19	WVLFIACAITSHES-----LTKCQOLQANSANSLIGTYVPOCKENGEEFKCQMGSTGCVWCGVYDKEILOTKRKGSPCSRKALPLIC	70
Db	171	WLLFEMSKNSLKEKOPTQTOPPKVLTCKQEVSHIPDWHGAFRPKCDENGNYIMPLQHGS 230	230
Qy	71	TGWCWCVDEQDEKIELTKTIGSPCCS 96	
Db	231	TGWCWCVFPNQTEVPHKTSRGRHNCs 256	256

RESULT 12
 NMHUND
 nitrogen precursor - human
 N; Alternative names: enactin,lin
 C; Species: Homo sapiens (man)
 C; Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Oct-2000
 C; Accession: A33327; A32437; A61367
 R; Nayachii, T.; Sanborn, D.; Hickok, N.J.; Olsen, D.R.; Fazio, M.J.; Chu, M.L.; Kr

DNA_8	581-594	198	RESULT 13
A;Title:	Human nidogen: complete amino acid sequence and structural domains deduced from A;Reference number: A33322; MUID:90091745		A;Protein name: protein 41 - human (fragment)
A;Accession:	A33322		C;Species: Homo sapiens (man)
A;Residues:	1-1247 <WAG>		C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 23-Jul-1999
A;Cross-references:	EMBL:M30269		C;Accession: 170687
R;Olson, D.R.; Nagayoshi, T.; Fazio, M.; Mattei, M.G.; Passage, E.; Weil, D.; Timpl, R.; Am. J. Hum. Genet. 44: 876-885, 1989		R;O'Sullivan, D.M.; Noonan, D.; Quaranta, V.	
A;Title:	Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1. Invest. Dermatol. 97, 281-285, 1991		J. Exp. Med. 166, 444-460, 1987
A;Reference number:	A61367; MUID:91302882		A;Title: Four La invariant chain forms derive from a single gene by alternate splicing
A;Accession:	A32437		A;Reference number: 155633; MUID:87252440
A;Molecule type:	mRNA		A;Accession: 170687
A;Residues:	667-1247 <OLS>		A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Cross-references:	EMBL:M27445; PIDN:AAA57261.1; PID:9602467		A;Molecule type: mRNA
A;Note:	the authors translated the codon AAG for residue 965 as Cys		A;Residues: 1-71 <RES>
A;Title:	Human nidogen gene: structural and functional characterization of the 5'-flanking region. J. Invest. Dermatol. 97, 281-285, 1991		A;Cross-references: GB:M28590; NID:9292154; PIDN:AAA3596.1; PID:9292155
A;Reference number:	A61367; MUID:91302882		C;Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin
A;Map position:	1q43-1q43		F;8-66/Domain: thyroglobulin type I repeat homology <THY1>
C;Superfamily:	nidogen; EGF homology; LDL receptor YWTD-containing repeat homology; thymosin beta-4; basement membrane; beta-hydroxyasparagine; calcium binding; cell binding; cd44		
C;Keywords:	basement membrane; beta-hydroxyasparagine; calcium binding; cell binding; cd44		
F;1-128/Domain:	signal sequence #status predicted <SIG>		
F;1-28/Domain:	nidogen #status predicted <MAT>		
F;1-29-1247/Domain:	EGF homology <EG1>		
F;1-67-708/Domain:	EGF homology <EG2>		
F;714-750/Domain:	cell attachment (R-G-D) motif		
F;762-800/Domain:	EGF homology <EG3>		
F;806-839/Domain:	EGF homology <EG4>		
F;849-919/Domain:	thryoglobulin type I repeat homology <THY1>		
F;910-1032/Domain:	LDL receptor YWTD-containing repeat homology <YWL>		
F;1033-1075/Domain:	LDL receptor YWTD-containing repeat homology <YW2>		
F;1076-1120/Domain:	LDL receptor YWTD-containing repeat homology <YW3>		
F;1121-1160/Domain:	LDL receptor YWTD-containing repeat homology <YW4>		
F;1161-1197/Domain:	LDL receptor YWTD-containing repeat homology <YW5>		
F;1212-1243/Domain:	EGF homology <EG6>		
F;1289-296/Binding site:	sulfate (Tyr) (covalent)		
F;729,819/Modified site:	erythro-beta-hydroxyasparagine (Asn) #status predicted		
F;756/Cross-link:	isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status predicted		
F;1137/Binding site:	carbohydrate (Asn) (covalent) #status predicted		
Query Match	11.7%	Score 150;	DB 2;
Best Local Similarity	42.9%	Pred. No. 9.5e-06;	Length 71;
Matches	27;	Mismatches	24;
Indels	0;	Gaps	0;
Db	65 NCS 67		
RESULT 14			
HLHMSG			
Class II histocompatibility antigen-associated gamma chain II, 41K splice form - murine			
N;Alternate names:	Ia-associated invariant chain II; invariant-chain proteolytic		
N;Contains:	class II histocompatibility antigen-associated gamma chain II, 31K sp		
C;Species:	Mus musculus (house mouse)		
C;Date:	17-Mar-1987 #sequence_revision 12-Apr-1996 #text_change 22-Jun-1999		
C;Accession:	B27866; A27866; A02244; A43530; S05099; S09284; A28956		
R;Koch, N.; Lauer, W.; Dohberstein, B.			
EMBO J. 6, 1671-1683, 1987			
A;Title:	primary structure of the gene for the murine Ia antigen-associated		
A;Reference number:	A91071; MUID:87275861		
A;Accession:	B27866		
A;Molecule type:	DNA		
A;Residues:	1-279 <KOC>		
A;Experimental source:	strain AKR		
A;Note:	the authors translated the codon AAC for residue 70 as Asp		
A;Accession:	A27866		
A;Residues:	1-191,256-279 <K02>		
A;Molecule type:	DNA		
A;Experimental source:	strain AKR		
A;Note:	the authors translated the codon AAC for residue 70 as Asp		
R;Singer, P.A.; Lauer, W.; Dembic, Z.; Mayer, W.E.; Lipp, J.; Koch, N.; Hammerling, G.I.; EMBO J. 3, 873-877, 1984			
A;Title:	Structure of the murine Ia-associated invariant (II) chain as deduced from the nucleotide sequence of the gene		
A;Reference number:	A02244; MUID:84207946		
A;Accession:	A02244		
A;Molecule type:	mRNA		
A;Residues:	1-191,256-279 <SIN>		
A;Cross-references:	GB:X0496; NID:953102; PIDN:CAA25191.1; PID:953103		
R;Eades, A.M.; Littfin, M.; Rahmsdorf, H.J.			
J. Immunol. 144: 4399-4409, 1990			
A;Title:	The IgM-gamma response of the murine invariant chain gene is mediated by		
A;Reference number:	A43530		
A;Status:	preliminary		
A;Molecule type:	DNA		
A;Residues:	1-25 <EAD>		

Submitted to the EMBL Data Library, March 1988
 A; Reference number: S03099
 A; Accession: S03099
 A; Molecule type: mRNA
 A; Residues: 1-58 <STO>
 A; Cross-references: EMBL:X07129; NID:952637; PID:CAA30141.1; PID:952638
 R; Zhi, L.; Jones, P.P.
 R; Miller, J.; Hatch, J.A.; Simonis, S.; Cullen, S.E.
 Nucleic Acids Res. 17, 447-448, 1989
 A; Title: Complete sequence of the murine invariant chain (II) gene.
 A; Reference number: S09284; MUID:8909846
 A; Accession: S09284
 A; Molecule type: DNA
 A; Residues: 1-9, 'M' 11-112, 'O' 114-228, 'STG' 232-279 <ZHU>
 A; Residues: 1-9, 'M' 11-112, 'O' 114-228, 'STG' 232-279 <ZHU>
 A; Prok. Natl. Acad. Sci. U.S.A. 85, 1359-1363, 1988
 A; Title: Identification of the glycosaminoglycan-attachment site of mouse invariant-chain (II) gene.
 A; Reference number: A28956; MUID:8B144436
 A; Accession: A28956
 A; Molecule type: mRNA
 A; Residues: 261-270 <MTI>
 C; Genetics:
 A; Introns: 26/2; 83/1; 109/3; 130/3; 161/3; 192/1; 256/1; 277/1
 C; Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin
 C; Keywords: alternative splicing; glycoprotein; transmembrane protein
 F; 1-29/Product: class II histocompatibility antigen-associated gamma chain II, 41K spl.
 F; 1-191, 256-279/Product: class II histocompatibility antigen-associated gamma chain II,
 F; 1-29/Domain: intracellular #status predicted <CYT>
 F; 196-254/Domain: extracellular #status Predicted <EXT>
 F; 113-119, 239, 253/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F; 265/Binding site: carbohydrate (ser) (covalent) #status experimental.

Query Match 10.5%; Score 135; DB 1; Length 279;
 Best Local Similarity 37.2%; Pred. No. 0.00054;
 Matches 32; Conservative 11; Mismatches 35; Indels 8; Gaps 2;
 C; Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Dec-1999
 C; Accession: A45403
 R; Moser, D.R.; Lowe Jr., W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, I.
 Mol. Endocrinol. 6, 1805-1814, 1992
 A; Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.
 A; Reference number: A45403; MUID:93125553
 A; Accession: A45403
 A; Status: preliminary
 A; Molecule type: mRNA; protein
 A; Residues: 1-258 <MOS>
 A; Cross-references: GB:S52770; NID:9263303; PID:AB24873.1; PID:9263304
 A; Experimental source: pulmonary artery endothelial cells
 A; Note: sequence extracted from NCBI backbone (NCBInr:122183; NCBP:122184)
 C; Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat
 F; 174-249/Domain: thyroglobulin type I repeat homology <THY>

Mon Mar 11 08:00:36 2002

us-09-445-480a-2.rpr

OM protein - protein search, using sw model													
Run on:													
March 8, 2002, 09:21:32 ; Search time 12.57 Seconds													
{without alignments}													
413.545 Million cell updates/sec													
Title:	US-09-445-480A-2												
Perfect score:	1287												
Sequence:	1 MAISQNQAKFSKGFFVVMWV.....VDEGGVVKPGSDVRFKRPTC 231												
Scoring table:	BLOSUM62												
Searched:	Gapop 10.0 , Gapext 0.5												
Total number of hits satisfying chosen parameters:	212252												
Minimum DB seq length:	0												
Maximum DB seq length:	200000000												
Post-processing:	Minimum Match 0%												
Listing first 45 summaries													
Database :	Issued_Patents_AA:*												
1:	/cgn2_6_ptodata/2/1aa/5A__COMB.pep:*												
2:	/cgn2_6_ptodata/2/1aa/5B__COMB.pep:*												
3:	/cgn2_6_ptodata/2/1aa/6A__COMB.pep:*												
4:	/cgn2_6_ptodata/2/1aa/6B__COMB.pep:*												
5:	/cgn2_6_ptodata/2/1aa/PCQSUS__COMB.pep:*												
Pred.	No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.												
SUMMARIES													
Result No.	Score	Query Match Length	DB ID	Description									
1	122.5	9.5	186	1 US-08-698-551-6	Sequence 6, Appli								
2	122.5	9.5	186	2 US-08-602-228-6	Sequence 6, Appli								
3	122.5	9.5	186	2 US-08-649-341A-5	Sequence 6, Appli								
4	122.5	9.5	186	2 US-08-494-440B-5	Sequence 6, Appli								
5	122.5	9.5	186	2 US-08-833-032A-5	Sequence 6, Appli								
6	122.5	9.5	186	2 US-08-833-031A-5	Sequence 6, Appli								
7	122.5	9.5	186	2 US-08-833-031A-5	Sequence 6, Appli								
8	122.5	9.5	186	2 PCT-US95-12724-6	Sequence 6, Appli								
9	122.5	9.5	237	6 5212074-7	Patent No. 5212074								
10	122.5	9.5	272	1 US-08-698-551-8	Sequence 8, Appli								
11	122.5	9.5	272	2 US-08-602-228-8	Sequence 8, Appli								
12	122.5	9.5	272	2 US-08-649-341A-8	Sequence 8, Appli								
13	122.5	9.5	272	2 US-08-494-440B-8	Sequence 8, Appli								
14	122.5	9.5	272	2 US-08-833-901B-8	Sequence 8, Appli								
15	122.5	9.5	272	2 US-08-839-032A-8	Sequence 8, Appli								
16	122.5	9.5	272	2 US-08-839-031A-8	Sequence 8, Appli								
17	122.5	9.5	272	5 PCT-US95-12724-8	Sequence 8, Appli								
18	122.5	9.5	272	5 PCT-US93-01196-3	Sequence 8, Appli								
19	118.5	9.2	271	5 5212074-6	Patent No. 5212074								
20	113.5	8.8	213	6 5212074-6	Patent No. 5212074								
21	113.5	8.8	322	6 5212074-1	Patent No. 5212074								
22	112	8.7	1196	1 US-08-444-121-4	Sequence 4, Appli								
23	112	8.7	1196	2 US-08-735-893-4	Sequence 4, Appli								
24	109.5	8.5	259	6 5212074-3	Patent No. 5212074								
25	109.5	8.5	4544	1 US-08-469-486-5	Sequence 52, Appli								
26	109.5	8.5	4544	2 US-08-659-551-5	Sequence 52, Appli								
27	107	8.3	85	3 US-08-604-965E-9	Sequence 9, Appli								
ALIGNMENTS													
RESULT	1	Query Match	9.5%	Score 122.5;									
US-08-698-551-6	Sequence 6, Application	US/0869851											
Patent No. 5212074	GENERAL INFORMATION:												
	APPLICANT: Lin, Lih-Ling												
	APPLICANT: Chen, Jennifer H.												
	APPLICANT: Schiavelli, Andrea												
	APPLICANT: Graham, James												
	APPLICANT: Chen, Lih-Ling												
	APPLICANT: Chen, Jennifer H.												
	APPLICANT: Schiavelli, Andrea												
	APPLICANT: Graham, James												
	APPLICANT: Chen, Lih-Ling												
	APPLICANT: Chen, Jennifer H.												
	APPLICANT: Schiavelli, Andrea												
	NAME: Brown, Scott A.												
	REGISTRATION NUMBER: 32,724												
	REFERENCE/DOCKET NUMBER: G15232D												
	TELECOMMUNICATION INFORMATION:												
	TELEPHONE: (617) 498-8224												
	FAX: (617) 876-5851												
	INFORMATION FOR SEQ ID NO: 6:												
	SEQUENCE CHARACTERISTICS:												
	LENGTH: 186 amino acids												
	TYPE: amino acid												
	TOPOLOGY: linear												
	MOLECULE TYPE: protein												
QY	151	ELEGTRROGRPCTERHUSEEARIKAHNSLRLVERMHE											
QY	151	ELEGTRROGRPCTERHUSEEARIKAHNSLRLVERMHE											
QY	151	: : : : : : : :											

Sequence 3, Appl
Sequence 4, Appl
Sequence 45, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 19, Appl
Sequence 7, Appl
Sequence 7, Appl
Patient No. 5512074
Patent No. 5525827
Sequence 34, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl

RESULT 2
US-08-602-228-6
Sequence 6, Application US/08602228
Patient No. 584365
GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavello, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649, 341A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI5232-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602, 228
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI5232C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-602-228-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

QY 151 ELEGTROQRPTCERHLSTCEEARIAHNSNLSRVEMVPECLLEDGSYNPVCWPS---T 206
Db 96 EMRQESEQG-PCCRHM-EASLQELKASPRMVPRAVYLPCDRKGFYKRKOCKPSGRKR 152
QY 207 GYCWCVDGGKVPGSD 223
Db 153 GICWCVDKYGMKLPGME 169

RESULT 4
US-08-494-440B-6
Sequence 6, Application US/08494440B
Patient No. 5849501
GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavello, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494, 440B
FILING DATE:
CLASSIFICATION: 435

GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS

RESULT 3
US-08-649-341A-6
Sequence 6, Application US/08649341A
Patient No. 5847059
GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A,

REGISTRATION NUMBER: 32,724

REFERENCE/DOCKET NUMBER: GI15232A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 186 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-494-440B-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;
 best Local Similarity 35.1%; Pred. No. 3.4e-05;
 Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRROGPRCERHSECEARIKAHNSLRLVEMFVPECLDGSTNPVCWPS---T 206
 Db 96 EMRQESEOG--PCCRHM-EASLQELKASPRMVPRAVYLPCDRKGFTKRKOCKPSRKRR 152

Qy 207 GYCWCWDEGGVKGPGSD 223
 Db 153 GICWCVDKYGMKLPGME 169

RESULT 5

US-08-533-901B-6

Sequence 6, Application US/08533901B

; Patent No. 5891675

GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling

APPLICANT: Chen, Jennifer H.

APPLICANT: Schiavella, Andrea

APPLICANT: Graham, James

TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patientin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/839,032A

CLASSIFICATION:

536

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A,

REGISTRATION NUMBER: 32,724

REFERENCE/DOCKET NUMBER: GI1232DDIV

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 186 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-839-032A-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;

best Local Similarity 35.1%; Pred. No. 3.4e-05;

Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRROGPRCERHSECEARIKAHNSLRLVEMFVPECLDGSTNPVCWPS---T 206
 Db 96 EMRQESEOG--PCCRHM-EASLQELKASPRMVPRAVYLPCDRKGFTKRKOCKPSRKRR 152Qy 207 GYCWCWDEGGVKGPGSD 223
 Db 153 GICWCVDKYGMKLPGME 169

RESULT 7

US-08-839-031A-6

Sequence 6, Application US/08839031A

; Patent No. 5948638

RESULT 10
US-08-698-551-8
; Sequence 8, Application US/08698551
; Patent No. 5712381
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievelbein, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602-228
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, Scott A.
; REFERENCE/DOCKET NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; APPLICATION NUMBER: US/08/698, 551
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A,
; REFERENCE/DOCKET NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TOPOLogy: Linear
; MOLECULE TYPE: protein
; US-08-698-551-8

Query Match 9 5%; Score 122.5; DB 1; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;
Oy 151 ELEGTROQGRPTCERHLSCEEARIKAHNSLRYEMFVPECLGDSYNNQWCAPS----T 206
Db 182 EMQESEQG--PCRRHM-EASLOELKASPRMVPRAVYLPCNCDRKGYKRKOCKPSGRKR 238
Oy 207 GYFWCVDGGKVPGSD 223
Db 239 GICWCVDKYGMKLPGME 255

RESULT 12
US-08-649-341A-8
; Sequence 8, Application US/08649341A
; Patent No. 5847059
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649, 341A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, Scott A.
; REFERENCE/DOCKET NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TOPOLogy: Linear
; MOLECULE TYPE: protein
; US-08-602-228-8

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;
Oy 151 ELEGTROQGRPTCERHLSCEEARIKAHNSLRYEMFVPECLGDSYNNQWCAPS----T 206
Db 182 EMQESEQG--PCRRHM-EASLOELKASPRMVPRAVYLPCNCDRKGYKRKOCKPSGRKR 238
Oy 207 GYFWCVDGGKVPGSD 223
Db 239 GICWCVDKYGMKLPGME 255

RESULT 11
US-08-602-228-8
; Sequence 8, Application US/08602228
; Patent No. 5843675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievelbein, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.

; TELECOMMUNICATION INFORMATION:
; TELEFAX: (617) 986-8851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TOPOLogy: linear
; MOLECULE TYPE: protein
; US-08-649-341A-8

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRROGRPTCERHLSCEEARIKAHNSLREVMVPECLEGSNPNVPCWPS---T 206
Db 162 EMRQESEQG--PCCRHM-EASLOELKASPRMVPRAVYLPCDRKGFYKRKOCKPSRGKRR 238
Qy 207 GYCWCVDDEGGKVPGSD 223
Db 239 GICWCVDKYGMKLPGME 255

RESULT 14
US-08-533-901B-8
; Sequence 8, Application US/08533901B
; Patent No. 5852173
; GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533, 901B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI5232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-8851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLogy: linear
MOLECULE TYPE: protein
; US-08-533-901B-8

RESULT 15

US-08-839-032A-8
; Sequence 8, Application US/08839032A
; Patent No. 5891675
; GENERAL INFORMATION:

Qy 151 ELEGTRROGRPTCERHLSCEEARIKAHNSLREVMVPECLEGSNPNVPCWPS---T 206
Db 182 EMRQESEQG--PCCRHM-EASLOELKASPRMVPRAVYLPCDRKGFYKRKOCKPSRGKRR 238
Qy 207 GYCWCVDDEGGKVPGSD 223
Db 239 GICWCVDKYGMKLPGME 255

RESULT 13
US-08-494-440B-8
; Sequence 8, Application US/08494440B
; Patent No. 5849501
; GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533, 901B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI5232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-8851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLogy: linear
MOLECULE TYPE: protein
; US-08-494-440B-8

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRROGRPTCERHLSCEEARIKAHNSLREVMVPECLEGSNPNVPCWPS---T 206
Db 182 EMRQESEQG--PCCRHM-EASLOELKASPRMVPRAVYLPCDRKGFYKRKOCKPSRGKRR 238
Qy 207 GYCWCVDDEGGKVPGSD 223
Db 239 GICWCVDKYGMKLPGME 255

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05; Mismatches 31; Indels 7; Gaps 3;
; US-08-649-341A-8

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05; Mismatches 31; Indels 7; Gaps 3;
; US-08-839-032A-8

APPLICANT: Graham, James
 TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/839,032A

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A,

REGISTRATION NUMBER: 32,724

REFERENCE/DOCKET NUMBER: GI5232DDIV

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-8951

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 272 amino acids

TYPE: amino acid

TOPOLOGY: Linear

MOLECULE TYPE: protein

; US-08-839-032A-8

Query Match 9 5%; Score 122.5; DB 2; Length 272;
 Best Local Similarity 35.1%; Pred. No. 5.6e-05; Indels 7; Gaps 3;
 Matches 27; Conservative 12; Mismatches 31;

Qy	151	ELECTROQGRPTCEHLSCEEARIKAHNSLRLVEMFPECLEDGSINPVQWPS---T	206
Db	: : : : : : : : : : : : : :		
	182	EMRQESEQQ--PCRHM-EASLQRKASPRMVRAYVAPNCDRKGFRRKOCKPSRGRR	238
Qy	207	GYCWCVDEGGYKVPSD	223
Db	239	GICWCVVDKYGMKLRGME	255

Search completed: March 8, 2002, 09:24:47
 Job time: 195 sec



Copyright (c) 1993 - 2000 Compugen Ltd.

OM Protein - protein search, using sw model

Run on: March 8, 2002, 09:21:07 ; Search time 23.39 seconds
 731.549 Million cell updates/sec

Title: US-09-445-480A-2
Perfect score: 1287
Sequence: 1 MALSONQAKFSKGFVVMIWV..... VDBGGVKVPGSDVRFKRPTC 231

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext: 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :

1: A_Geneseq_1101.*
 2: /SIDS8/gcdata/geneseq/geneseq/AA1980.DAT: *
 3: /SIDS8/gcdata/geneseq/geneseq/AA1982.DAT: *
 4: /SIDS8/gcdata/geneseq/geneseq/AA1983.DAT: *
 5: /SIDS8/gcdata/geneseq/geneseq/AA1984.DAT: *
 6: /SIDS8/gcdata/geneseq/geneseq/AA1985.DAT: *
 7: /SIDS8/gcdata/geneseq/geneseq/AA1987.DAT: *
 8: /SIDS8/gcdata/geneseq/geneseq/AA1988.DAT: *
 9: /SIDS8/gcdata/geneseq/geneseq/AA1989.DAT: *
 10: /SIDS8/gcdata/geneseq/geneseq/AA1990.DAT: *
 11: /SIDS8/gcdata/geneseq/geneseq/AA1991.DAT: *
 12: /SIDS8/gcdata/geneseq/geneseq/AA1992.DAT: *
 13: /SIDS8/gcdata/geneseq/geneseq/AA1993.DAT: *
 14: /SIDS8/gcdata/geneseq/geneseq/AA1994.DAT: *
 15: /SIDS8/gcdata/geneseq/geneseq/AA1995.DAT: *
 16: /SIDS8/gcdata/geneseq/geneseq/AA1996.DAT: *
 17: /SIDS8/gcdata/geneseq/geneseq/AA1997.DAT: *
 18: /SIDS8/gcdata/geneseq/geneseq/AA1998.DAT: *
 19: /SIDS8/gcdata/geneseq/geneseq/AA1999.DAT: *
 20: /SIDS8/gcdata/geneseq/geneseq/AA2000.DAT: *
 21: /SIDS8/gcdata/geneseq/geneseq/AA2001.DAT: *
 22: /SIDS8/gcdata/geneseq/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1287	100.0	231	Actinia equina L.
2	259.5	20.2	447	Actinia equina L. Mouse TANGO 209 po
3	259.5	20.2	447	Actinia equina L. Mouse TANGO 209 va
4	259.5	20.2	447	Actinia equina L. Mouse TANGO 209 va
5	259.5	20.2	447	Actinia equina L. Mouse TANGO 209 va
6	254.5	19.8	446	Actinia equina L. Human TANGO 209 va
7	254.5	19.8	446	Actinia equina L. Human TANGO 209 va
8	254.5	19.8	446	Actinia equina L. Human TANGO 209 va
9	254.5	19.8	446	Actinia equina L. Human TANGO 209 va
10	252.5	19.6	435	Actinia equina L. Human TANGO 209 va
11	247.5	19.2	238	Actinia equina L. Human prostate can

RESULT 1
 ID AAW87988 standard; Protein; 231 AA.
 AC AAW87988;
 XX DT 09-APR-1999 (first entry)
 DE Actinia equina L. equistatin protein.
 XX KW Actinia equina L.; equistatin; protection; insect; nematode; plant;
 KW I repeated thyroglobulin domain; cysteine protease inhibitor;
 KW aspartic protease inhibitor; sea anemone.
 OS Actinia equina.
 XX FH Key Location/Qualifiers
 FT Peptide 32..1
 FT protein /label= signal
 PN W0985068-A2.
 XX PD 23-DEC-1998.
 XX PF 18-JUN-1998; 98WO-NL00352.
 XX PR 18-JUN-1997; 97EP-020177.
 XX (CPRO-) CPR0-DLO CENT PLANTENVERDELINGS REPROD.
 XX PI Bosch HJ, Gruden K, Jongasma MA, Ienarcic B, Stiekema WJ;
 PI Strukelj B, Turk V;
 XX

Query Match 100.0%; Score 1287; DB 20; Length 231;
 Best Local Similarity 100.0%; Pred. No. 4 1e-104; Mismatches 0; Indels 0; Gaps 0;

CC 1 MALSONAKSKSGVNMWFLIACATTSPEASLTKRQOLQASANSGLIGTYPOCKETG 60
 CC 1 malisqnaksksgtvvmwlviaatiteasltkceqiqgasansgligtypketg 60

CC 61 REPEKCGSGYWCWVDEGSEKLGITKRGSPCSRKAALTCOMMOAIVNPWG 120
 CC 61 epekcgwsgtgcwvdeogkeilgtkgirkspacsrkaaltcgmmalivnpwg 120

CC 121 PRSKCKADGSDIDEVOCASNGEYCCKKKKELEGTRGROGRCTERHLSCEEERIKHSN 180
 CC 121 pssckadgsiddevocasngcycckkkkelegtrgrotgrctehlsceearikhsn 180

CC 181 SIRVEMEVPSCLGDSYMPQWPGYCWVBRGGKVPSPDVKRPTC 231
 CC 181 sirvenevpsclgdsympqwpstgycwcvdegkvkvpqsdvfkptc 231

RESULT 2

Db AAB48111 standard; Protein: 447 AA.

XX AAB48111;

AC 02-APR-2001 (first entry)

DT DE Mouse TANGO 209 polypeptide.

XX X TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse; transmembrane protein; antiarrhythmic; cerebroprotective; arteriosclerosis; antiasthmatic; neuroprotective; cytostatic; cardiotropic; hepatotropic; antiinflammatory; antidiabetic; antiinfertility; antipyretic; vasotropic; analgesic; nephrotoxic; hemostatic; antilipemic; osteopathic; ophthalmological; antisickling; antiulcer; vulnerary.

XX OS Mis sp.

XX XX

PN WO2000059885-A2.

XX PD 23 NOV 2000.

XX PT 15-MAY-2000; 2000WO-US13361.

XX PR 14-MAY-1999; 99US-0312559.

XX PS Claim 15; Fig 1; 82pp; English.

PA (MILL-) MILLENIUM PHARM INC.

PI Pan Y, Leiby KR;

XX PT 2001-024999/03

DR N-PSDB; AAC84389, AAC84390.

XX PA Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas.

XX PS Claim 8; Fig 18A-E; 209pp; English.

CC The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain intracranial hemorrhage, pancreatitis, diabetes, angina, hypotension, heart disease, myocardial disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoparadrenism, hyperadrenalinism, cushing's syndrome, neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune Albright syndrome, infertility, uterine disorders, viral disease. The present sequence represents the mouse TANGO 209 polypeptide.

CC SQ Sequence 231 AA;

Query Match 100.0%; Score 1287; DB 20; Length 231;
 Best Local Similarity 100.0%; Pred. No. 4 1e-104; Mismatches 0; Indels 0; Gaps 0;

CC 41 QASANSLIGTYPOCKETGEEFKQCWSTGKWCWVDEOCKELGKRL-RQSPCSRKR 99
 CC 98 qeqarkkefgvfipecadgtysqvqchaytgcwvtpgrisgtavanhkptcp-- 154

CC 100 AALTICQMMQALIVNGWGCD-----PPSKCKADGSEFEVQCCASNGEYCIVVKKGK 151
 CC 155 -----gsinekvprqregakdappaletgpqqgediedasryptiwteyvksq 205

CC 152 LRGTRKOGAPTCRHLSECEARAKAHNSLRVEMVPELEODSYNVOWPSTGWC 211

CC 206 ntknksasscdhepsaleekapknnd---vvipecahgglykpvqchpstgycw 260

CC 212 V-DEGCKVPGSIVRFKPTC 231

CC 261 vlvdtgrgrpgtstryeqpkc 281

RESULT 3

AAV48142 ID AAB48142 standard; Protein: 447 AA.

XX OS AAB48142;

AC AC

XX
 XX 02-APR-2001 (first entry)
 DE
 XX
 DE MOUSE TANGO 209 variant 1 polypeptide.
 XX
 KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytostatic; cardiant; hepatotrophic;
 KW antiinflammatory; antidiabetic; antiinfertility; antipyretic; vasoactive;
 KW antirheumatic; nephrotoxic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antisickling; antiluicer; vulnerary; variant.
 OS Mus sp.
 FH
 FT Msc-difference 65 Location/Qualifiers
 FT /label = E65D
 FT /note= "wild-type Glu is replaced by Asp"
 XX
 PN WO200069885-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 15-MAY-2000; 2000WO-US13361.
 PR 14-MAY-1999; 99US-0312359.
 XX
 PA (NILL-) MILLENNIUM PHARM INC.
 XX
 PI Pan Y.; Leiby KR;
 XX
 DR WPI; 2001-024999/03.
 XX
 DR N-PSDB; AAC84410.
 XX
 PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g., cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 PS Claim 8; Page -; 209pp; English.
 XX
 CC The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding Secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, trophic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, Infarction, heart
 CC intracranial hemorrhage, pancreatitis, diabetes, andina hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoparadrenism, hyperadrenalinism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse TANGO 209 variant polypeptide.
 CC Note: the present sequence has been constructed using the
 CC information provided in the specification.
 XX
 SQ Sequence 447 AA;

Query Match 20.2%; Score 259.5; DB 22; Length 447;
 Best Local Similarity 28.9%; Pred. No. 1; 3e-14;
 Matches 58; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

QY 41 QASANSLGLITYVPOCKETGEFFERKQCWGSGTCWCVDGKELGTL1-RGSPDCCSRK 99

Db 98 qeqarkefqgvifipecdgqdglysqvqchsytgycwcvngprisgtavhktprcp-- 154
 QY 100 AALTICQMMQATIVNVGWCQ-----PPSKRADGSFDEVCCASNGECYCVDKRGKE 151
 Db 155 -----gsinekvqregakdadaapeltqpgdeediasyptlwteqksrq 205
 QY 152 LEGTROQRPTCERHLCEERAKHNSLRLVENFVPCLEDSYNQCPSTGWC 211
 Db 206 nktknasscdaehqsaaleakqkdn---wvpecahglyxpqvcphstycwc 260
 QY 212 V-DGGGVKPGSDVRFKRPTC 231
 Db 261 vivatgrpipgistryeqpk 281

RESULT 4
 AAB48143
 ID AAB48143 standard; Protein: 447 AA.
 XX
 AC AAB48143;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Mouse TANGO 209 variant 2 polypeptide.
 XX
 KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytostatic; cardiant; hepatotrophic;
 KW antiinflammatory; antidiabetic; antiinfertility; antipyretic; vasoactive;
 KW antirheumatic; nephrotoxic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antisickling; antiluicer; vulnerary; variant.
 OS Mus sp.
 FH
 FT Msc-difference 77 Location/Qualifiers
 FT /label = E77D
 FT /note= "wild-type Glu is replaced by Asp"
 XX
 PN WO200069885-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 15-MAY-2000; 2000WO-US13361.
 PR 14-MAY-1999; 99US-0312359.
 XX
 PA (NILL-) MILLENNIUM PHARM INC.
 XX
 DR N-PSDB; AAC84411.
 XX
 PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g., cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 PS Claim 8; Page -; 209pp; English.
 XX
 CC The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding Secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, trophic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, Infarction, heart
 CC intracranial hemorrhage, pancreatitis, diabetes, andina hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoparadrenism, hyperadrenalinism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse TANGO 209 variant polypeptide.
 CC Note: the present sequence has been constructed using the
 CC information provided in the specification.
 XX
 SQ Sequence 447 AA;

CC herniations, meningitis, ischemic brain or heart disease, infarction,
 CC intracranial hemorrhage, pancreaticitis, diabetes, angina, hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, hypertension's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalinism, hyperadrenalinism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse TANGO 209 variant polypeptide.
 CC Note: the present variant sequence has been constructed using the
 XX information provided in the specification.

SQ Sequence 447 AA;

Query Match Score 259.5; DB 22; Length 447;
 Best Local Similarity 28.9%; Pred. No. 1.3e-14; Mismatches 86; Indels 27; Gaps 5;
 Matches 58; Conservative 30; PSSM 95;

OY 41 QASANSGLIGTYVPOCKTGEFEERKQCWMSGTCWCVNEDGKGKIK-RGSPDCSRK 99
 Db 98 qeqarkefqgvfipecondgdtysqvchsytgycwcvtpngprisgtavahktprcp-- 154

OY 100 AALTLCOMMQAIIVNVPGCG-----PPSKAKDGFDEVQCCASNGECYCVDKKGKE 151
 Db 155 -----gsinekvparegagkadaaapaletqpgediasryptlwteqvkrsq 205

OY 152 LEGTROQGRPTCERHLSCEEARIKAHNSLRLVMFPECLEDGSYNPVQCPSTGWC 211

Db 206 nktknksasscdqehqsaileekapkndn-----vvipecahgglykpvcqchpstgycwc 260

OY 212 V-DEGGVKVPGSDYRFRKTC 231
 Db 261 vlvdtgrppgtstryeqpkc 281

RESULT 5

ID AAB48144; Standard; Protein; 447 AA.

XX AAB48144;
 AC AAB48144;
 DT 02-APR-2001 (first entry)
 DE Mouse TANGO 209 variant 3 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antiangiogenic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytostatic; cardiotropic; hepatoprotective;
 KW antiflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antirheumatic; nephroprotective; hemostatic; antilipidemic; osteopathic;
 KW ophthalmological; antisickling; antiulcer; vulnerary; variant.
 OS Mus sp.

XX Key difference 93
 FH Location/Qualifiers
 FT Misc-difference 93
 FT /label= E93D
 FT /note= "wild-type Glu is replaced by Asp"

Query Match Score 259.5; DB 22; Length 447;
 Best Local Similarity 28.9%; Pred. No. 1.3e-14; Mismatches 86; Indels 27; Gaps 5;
 Matches 58; Conservative 30; PSSM 95;

OY 41 QASANSGLIGTYVPOCKTGEFEERKQCWMSGTCWCVNEDGKGKIK-RGSPDCSRK 99
 Db 98 qeqarkefqgvfipecondgdtysqvchsytgycwcvtpngprisgtavahktprcp-- 154

OY 100 AALTLCOMMQAIIVNVPGCG-----PPSKAKDGFDEVQCCASNGECYCVDKKGKE 151
 Db 155 -----gsinekvparegagkadaaapaletqpgediasryptlwteqvkrsq 205

OY 152 LEGTROQGRPTCERHLSCEEARIKAHNSLRLVMFPECLEDGSYNPVQCPSTGWC 211

Db 206 nktknksasscdqehqsaileekapkndn-----vvipecahgglykpvcqchpstgycwc 260

OY 212 V-DEGGVKVPGSDYRFRKTC 231
 Db 261 vlvdtgrppgtstryeqpkc 281

RESULT 6

ID AAB48107; Standard; Protein; 446 AA.

XX AAB48107;
 AC AAB48107;
 DT 02-APR-2001 (first entry)
 DE Human TANGO 209 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antiangiogenic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytostatic; cardiotropic; hepatoprotective;
 KW antiflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antirheumatic; nephroprotective; hemostatic; antilipidemic; osteopathic;

XX Paul Y, Leiby KR;

Best Local Similarity		28.2%		Pred.		NO. 3.6e-14;			
Matches		59; Conservative		29; Mismatches		43; Gaps			
QY	41	QASANSGHLIGTYWPCKERGEEFRKQCMGSTCWCYDPEDEKELGKTI-RGSPDCSRK	99	CC	promote ulcer healing, treat bone cancer, achondroplasia, myeloma,	CC			
QY	98	qeqarkefqvifpcndgqysqvctgycvtprisgtavahktprc---	153	CC	osteoporosis, fibromatosis, dysplasia, scoliosis, osteoarthritis, brain	CC			
Db	100	AALTLCQMQATIVNPGWCGP--PSCKADGSDEV-----	144	CC	hematopoiesis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain	CC			
QY	154	-----pysmeklpqregtqktddaaapalaetqpgagdeediasryptlw	197	CC	intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart	CC			
QY	145	VDKKGKELEGTRQGRPCER-HLSECEEARIAHNSL-RVEMVPECTEDSYNPVCW	203	CC	disease, pulmonary heart disease, rheumatic fever, congenital heart	CC			
Db	198	teqvsqrnqtnksvsscdqehqsaleakqpkndn---vvipecahgglypvqch	252	CC	disease, myocardial disease, atherosclerosis, hypertension, jaundice,	CC			
QY	204	PSTGYCWCV-DEGGVKVPPGSDVRFKRPTC	231	CC	hepatitis, cirrhosis, glomerulonephritis, Goodpasture's syndrome,	CC			
Db	253	pstgycwcvlvdtgrgrpgtstryeqpkc	281	CC	sickle cell disease, renal failure, ischemic bowel disease, Crohn's	CC			
Note: the present variant sequence has been constructed using the information provided in the specification.									
SQ	Sequence								
446 AA;									
RESULT		8		Best Local Similarity		19.8%			
ID	RAB48140	standard; Protein: 446 AA.		Matches	59; Conservative	28.2%	Pred. NO. 3.6e-14;		
XX	XX	/label= "Wild-type Glu is replaced by Asp"		Mismatches	78;	DB	22; Length 446;		
AC	AAB48140;	/note= "Wild-type Glu is replaced by Asp"		Indels	43;	Gaps	7;		
DT	02-APR-2001	(first entry)							
DE	Human TANGO 209 variant 2 polypeptide.								
XX	XX	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;							
KW	transmembrane protein; antiangiogenic; arteriosclerosis; antidiabetic; neuroprotective; cytostatic; cardiotropic; hepatotropic; vasotrophic; antinflammatory; nephrotoxic; hemostatic; antilipemic; osteopathic; ophthalmological; antisickling; antiulcer; vulnerary; variant.								
KW	antidiabetic; antiinfectivity; antipyretic; vasotrophic; antinflammatory; nephrotoxic; hemostatic; antilipemic; osteopathic; ophthalmological; antisickling; antiulcer; vulnerary; variant.								
OS	OS	Homo sapiens.							
FH	Key	Location/Qualifiers							
FT	FT	Misc-difference 77							
FT	FT	/label= E77D							
PN	WO200069885-A2.	/note= "Wild-type Glu is replaced by Asp"							
PD	23-NOV-2000.								
XX	XX	PF 15-MAY-2000; 2000WO-US13361.							
PR	14-MAY-1999;	99US-0312359.							
PA	(MIL-L) MILLENNIUM PHARM INC.								
XX	XX	Pan Y, Leiby KR; Pan Y, Leiby KR; WPI: 2001-024969/03.							
DR	N-PSDB; AAC844408.								
XX	XX	Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -							
PS	PS	Claim 8; Page 7; 209PP; English.							
RESULT		9		Best Local Similarity		19.8%			
ID	AAB48141	standard; Protein: 446 AA.		Matches	59; Conservative	28.2%	Pred. NO. 3.6e-14;		
XX	XX	/label= "Wild-type Glu is replaced by Asp"		Mismatches	78;	DB	22; Length 446;		
AC	AAB48141;	/note= "Wild-type Glu is replaced by Asp"		Indels	43;	Gaps	7;		
DT	02-APR-2001	(first entry)							
DE	Human TANGO 209 variant 3 polypeptide.								
XX	XX	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;							
KW	transmembrane protein; antiangiogenic; arteriosclerosis; antidiabetic; neuroprotective; cytostatic; cardiotropic; hepatotropic; vasotrophic; antinflammatory; antisickling; antiulcer; vulnerary; variant.								
KW	antidiabetic; antiinfectivity; antipyretic; vasotrophic; antinflammatory; nephrotoxic; hemostatic; antilipemic; osteopathic; ophthalmological; antisickling; antiulcer; vulnerary; variant.								
OS	OS	Homo sapiens.							
FH	Key	Location/Qualifiers							
FT	FT	Misc-difference 93							
FT	FT	/label= E93D							
PN	WO200069885-A2.	/note= "Wild-type Glu is replaced by Asp"							
PD	23-NOV-2000.								
PR	15-MAY-2000; 2000WO-US13361.								
XX	14-MAY-1999;	99US-0312359.							

CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ3438, and AAY41685 to AAY1774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.

XX
SQ Sequence 436 AA:

Query Match 12.1%; Score 156; DB 20; Length 436;
Best Local Similarity 40.9%; Pred. No. 1.3e-05; Mismatches 25; Indels 2; Gaps 1;
Matches 27; Conservative 12; Mismatches 25; Indels 2; Gaps 1;

Qy 31 EASLTKQOOLQASANSGLIGYVPOCKENGEEFEERKOCWSTGYWCVCBDGKEILGTR 90
Db 318 qtelsniqkrqavkk-llqayipicdeddykptqchsgvqqcvcdrygnevngsrin 375

Qy 91 GSPDCS 96
| ||:
Db 376 gvdca 381

Search completed: March 8, 2002, 09:24:28
Job time: 201 sec

